

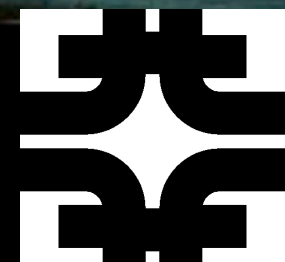
$D\bar{D}$ Higgs Combination and the $H \rightarrow \tau\tau$ analysis



Gabriel Facini
Northeastern University



August 30, 2011
SUSY@Fermilab



Standard Model Mass Generation



Standard Model (SM)

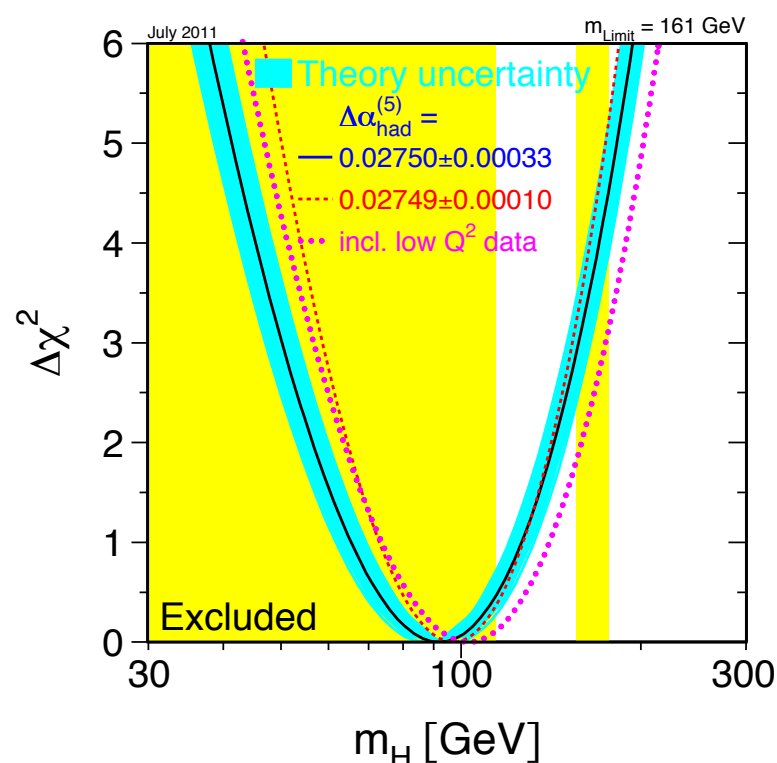
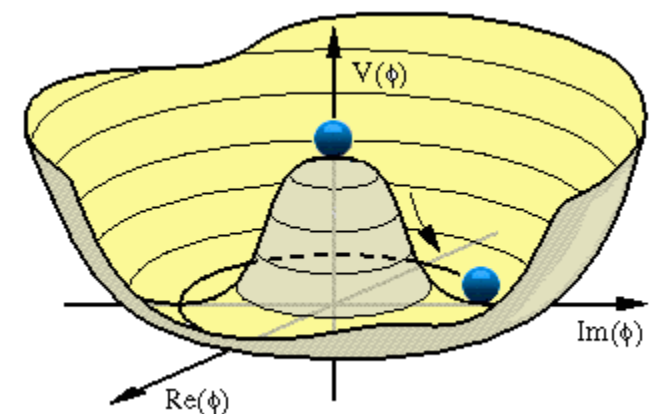
extremely successful as a predictive tool

without mass generation mechanism is still a **theory of massless particles**

Higgs field has a non-zero vacuum expectation value

Symmetry of potential is broken **spontaneously** to reach the ground state

Mass of the resultant Higgs particle is not predicted.

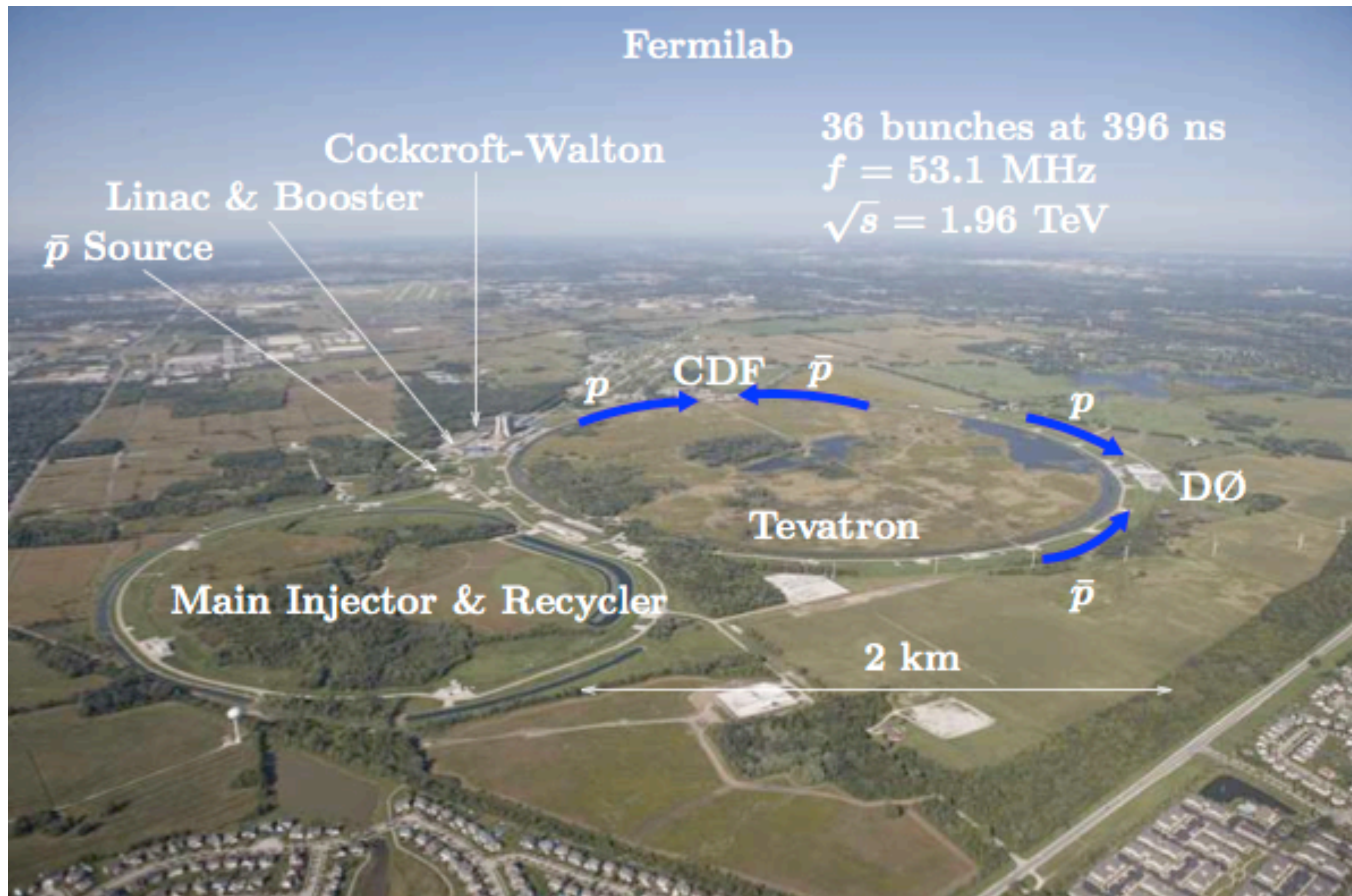


Theoretical constraints: $m_H < 161$ GeV at 95% CL.

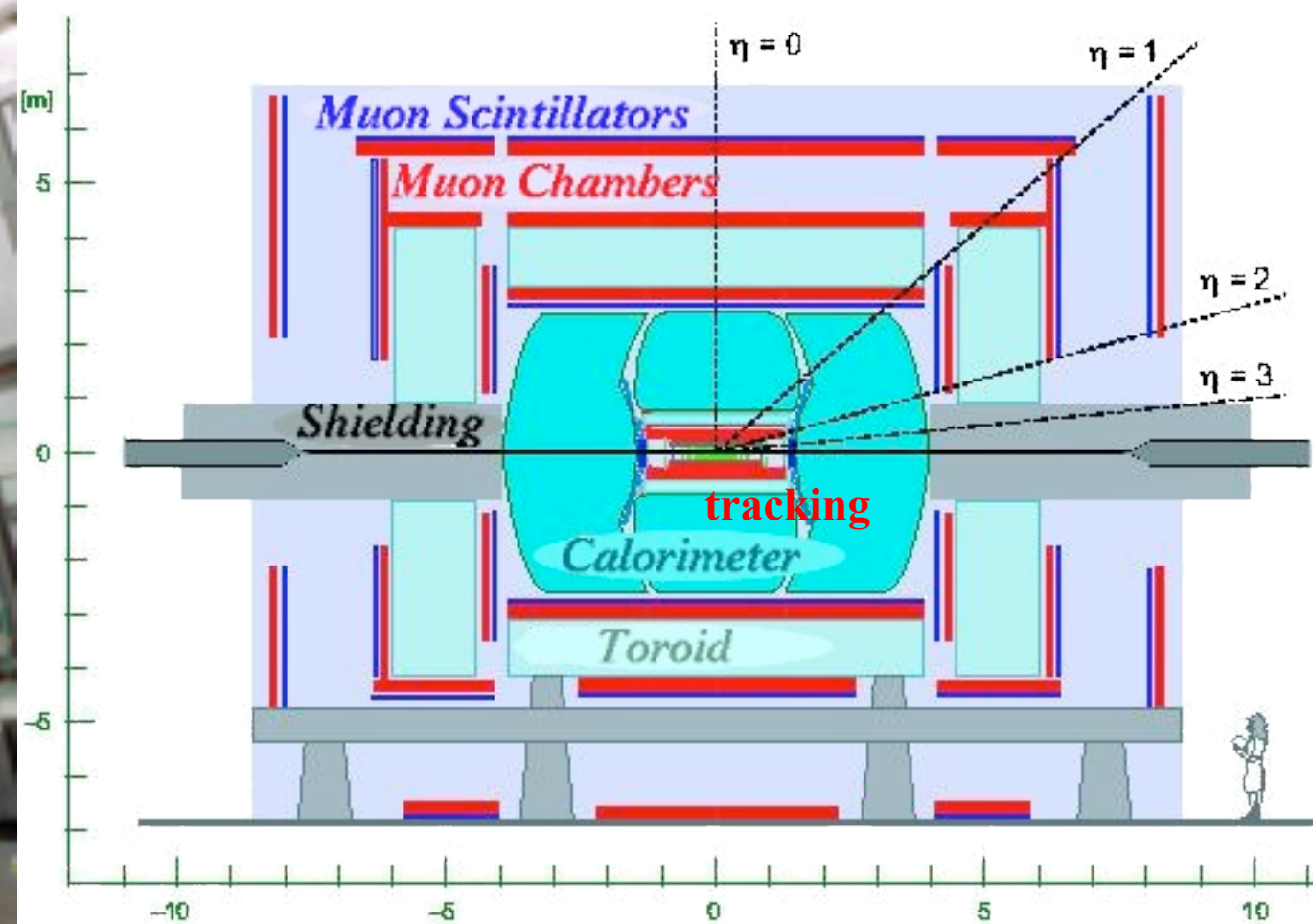
Direct searches: exclude $m_H < 114.4$ GeV at 95% CL

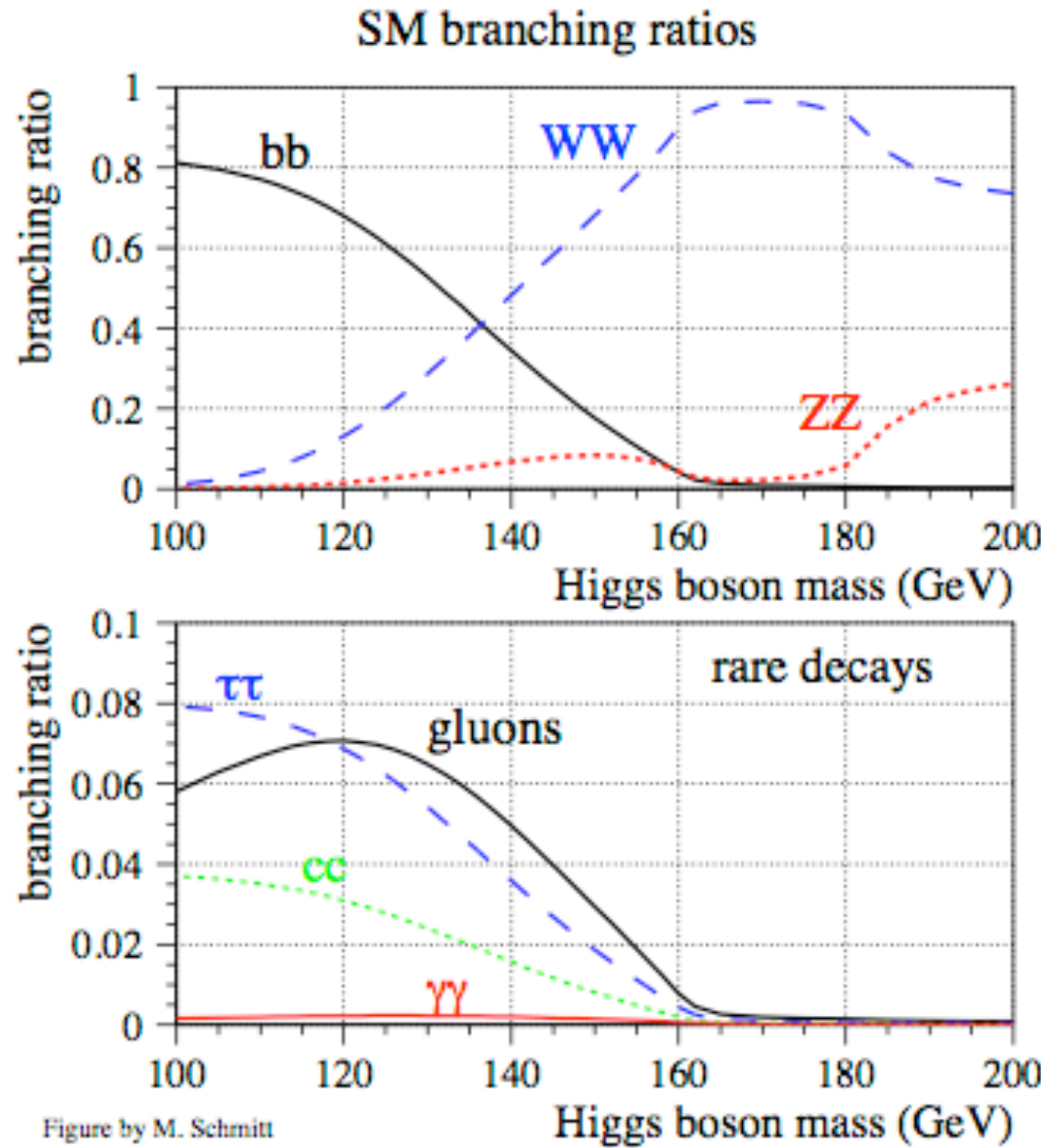
Tevatron and LHC have excluded a region ~ 165 GeV

The Tevatron



The DØ Detector

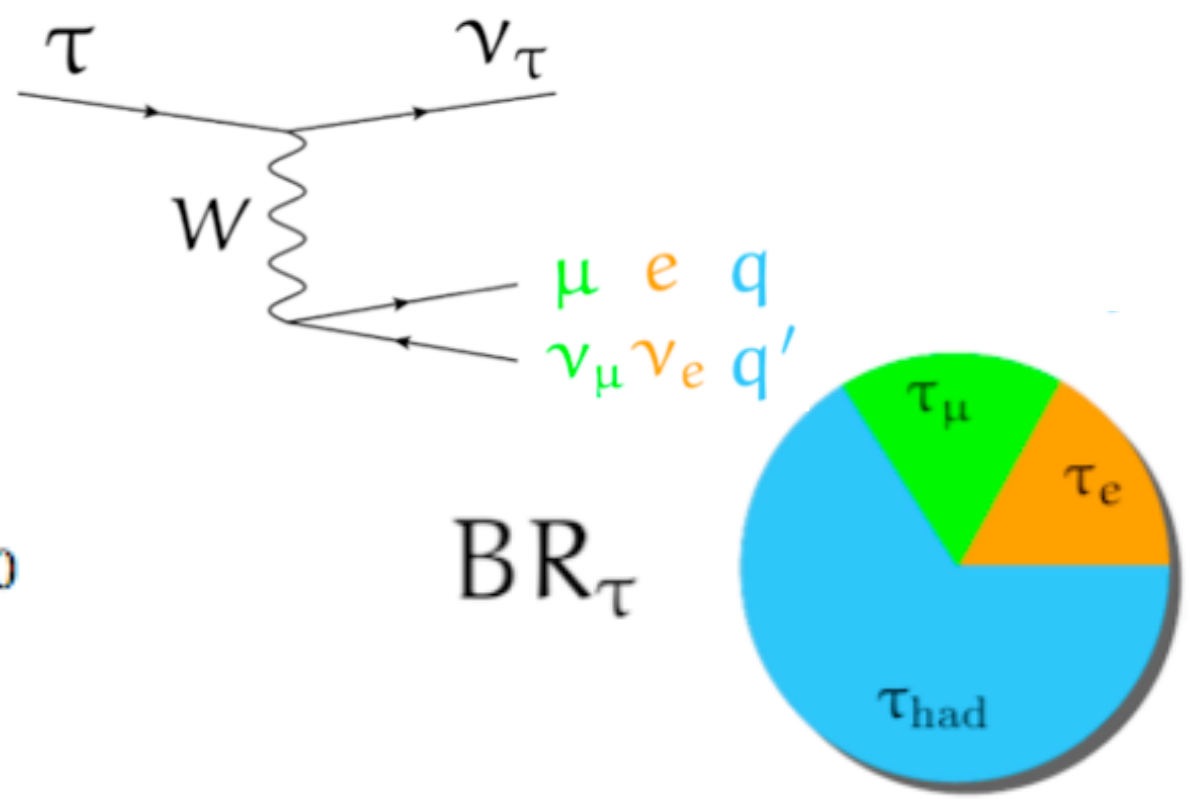




Range split into
 low mass : $H \rightarrow b\bar{b}$ mostly
 high mass : $H \rightarrow W^+W^-$ mostly

$H \rightarrow \gamma\gamma$ has a small BR but is a clean channel

Analysis involving τ decays can contribute across the entire mass range



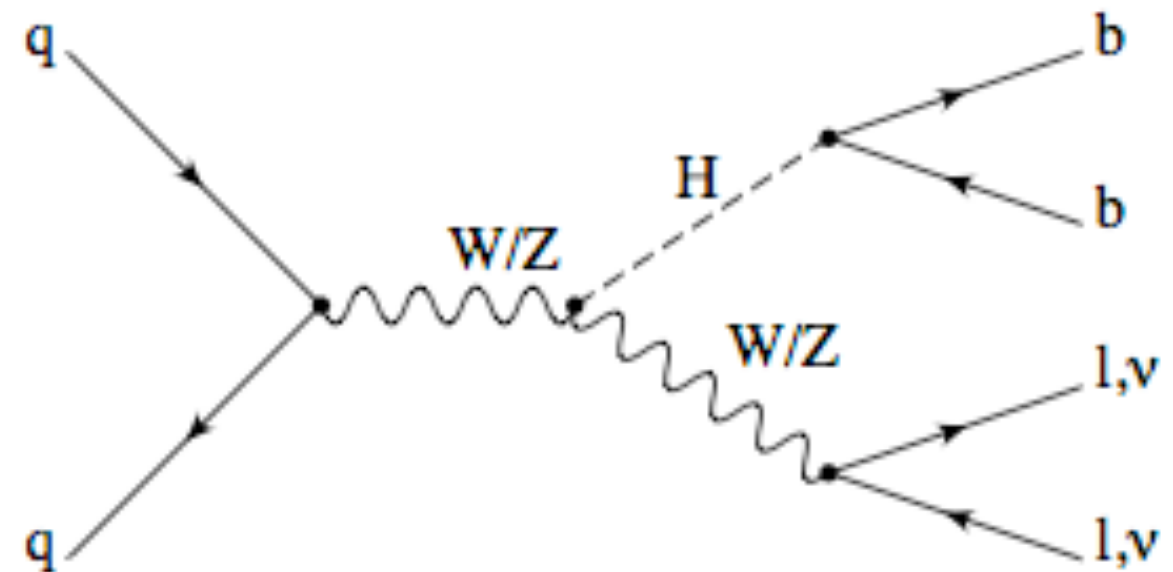
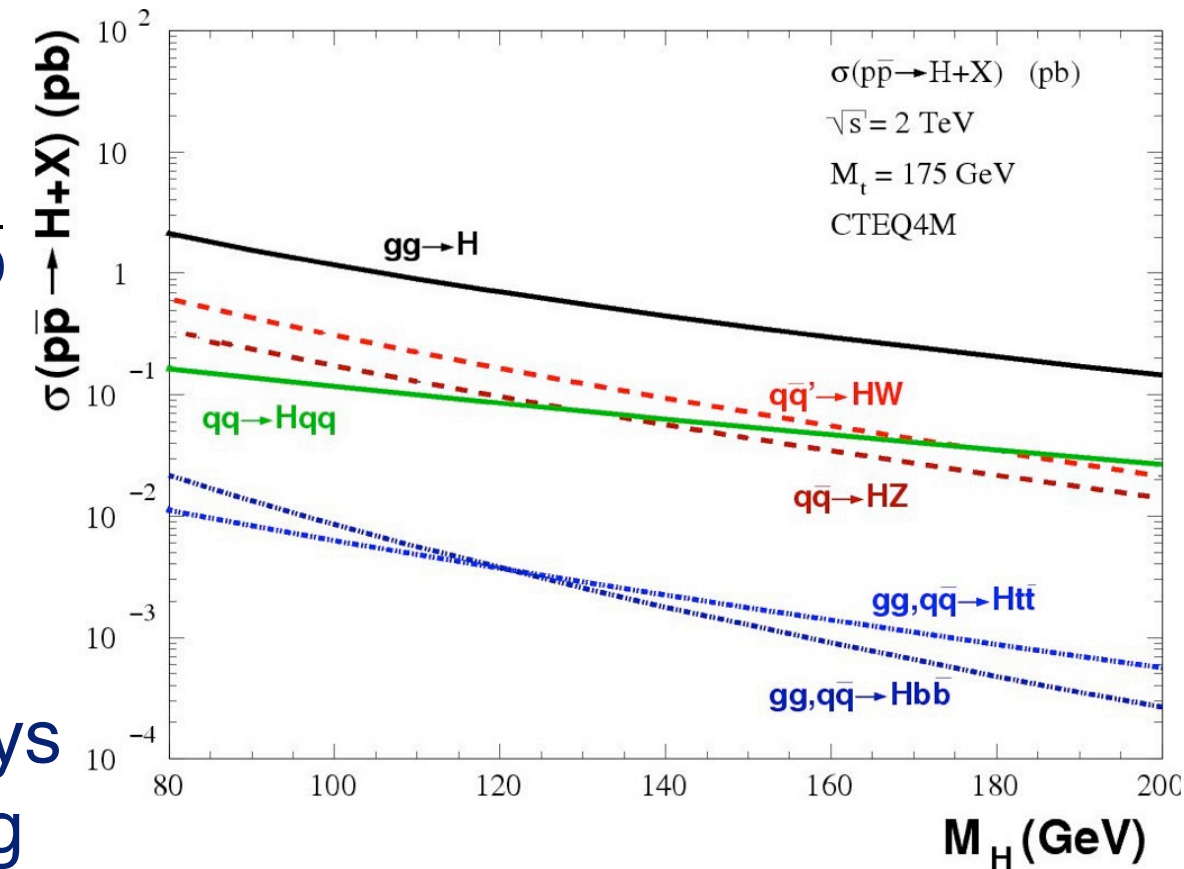
For $m_H < 135$ GeV the primary decay is $H \rightarrow b\bar{b}$

Primary production $gg \rightarrow H \rightarrow b\bar{b}$ but $q\bar{q} \rightarrow b\bar{b}$ background is **insurmountable**.

With $qq \rightarrow VH$ the gauge boson leptonic decays provide a handle to combat the overwhelming QCD background.

$$\begin{aligned} qq &\rightarrow W^\pm H \rightarrow l^\pm \nu b\bar{b} \\ qq &\rightarrow ZH \rightarrow \nu\nu b\bar{b} \\ qq &\rightarrow ZH \rightarrow l+l b\bar{b} \end{aligned}$$

Dominant contribution to the sensitivity.



Low Mass Approach



Major backgrounds are V +jets, $t\bar{t}$, and multijet production (MJ)

Control samples used to validate modeling

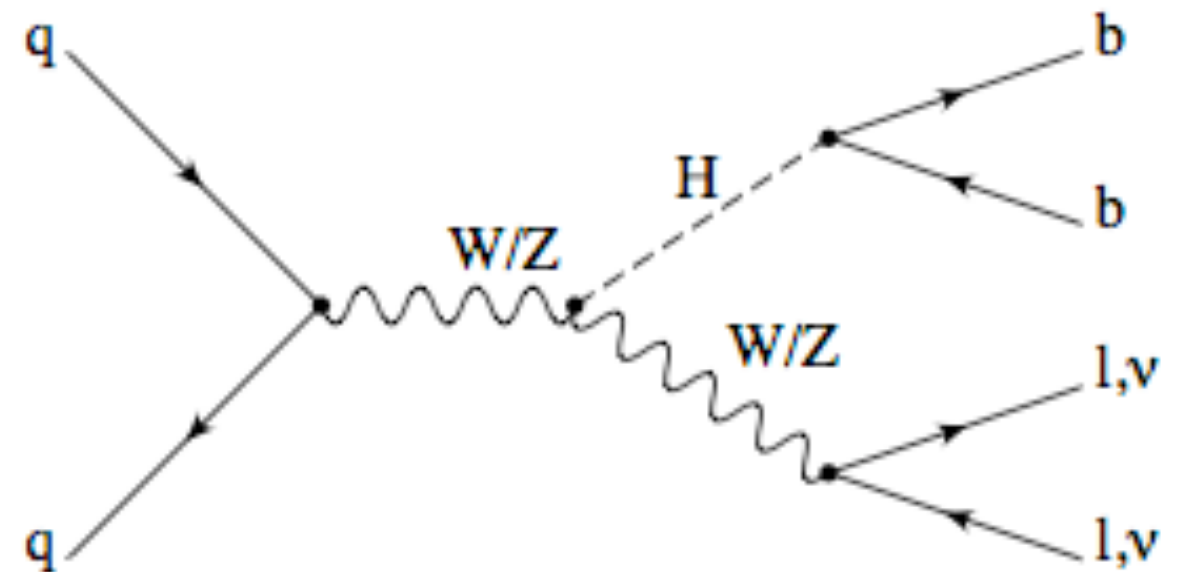
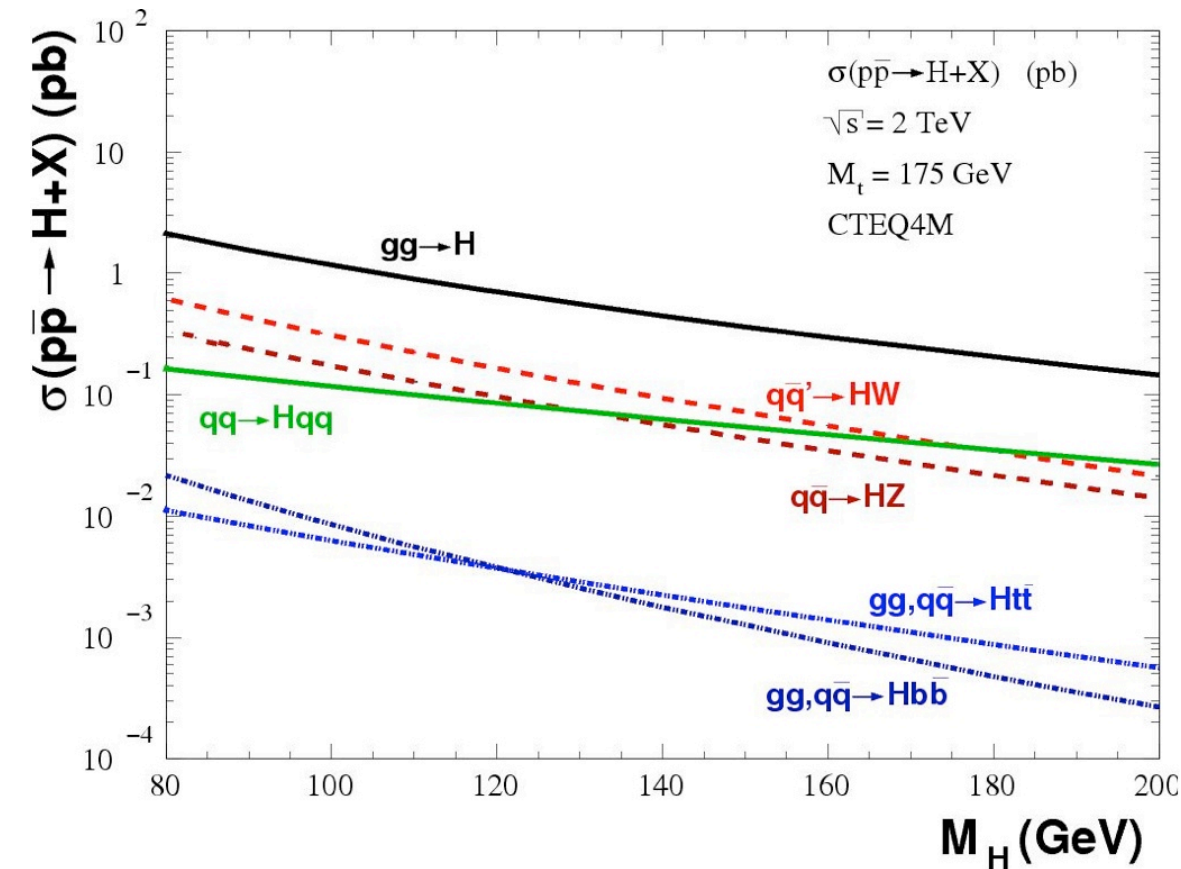
Multijet background - data driven method

b -jet identification - create orthogonal search channels

Use of multivariate discriminants

Remove MJ background

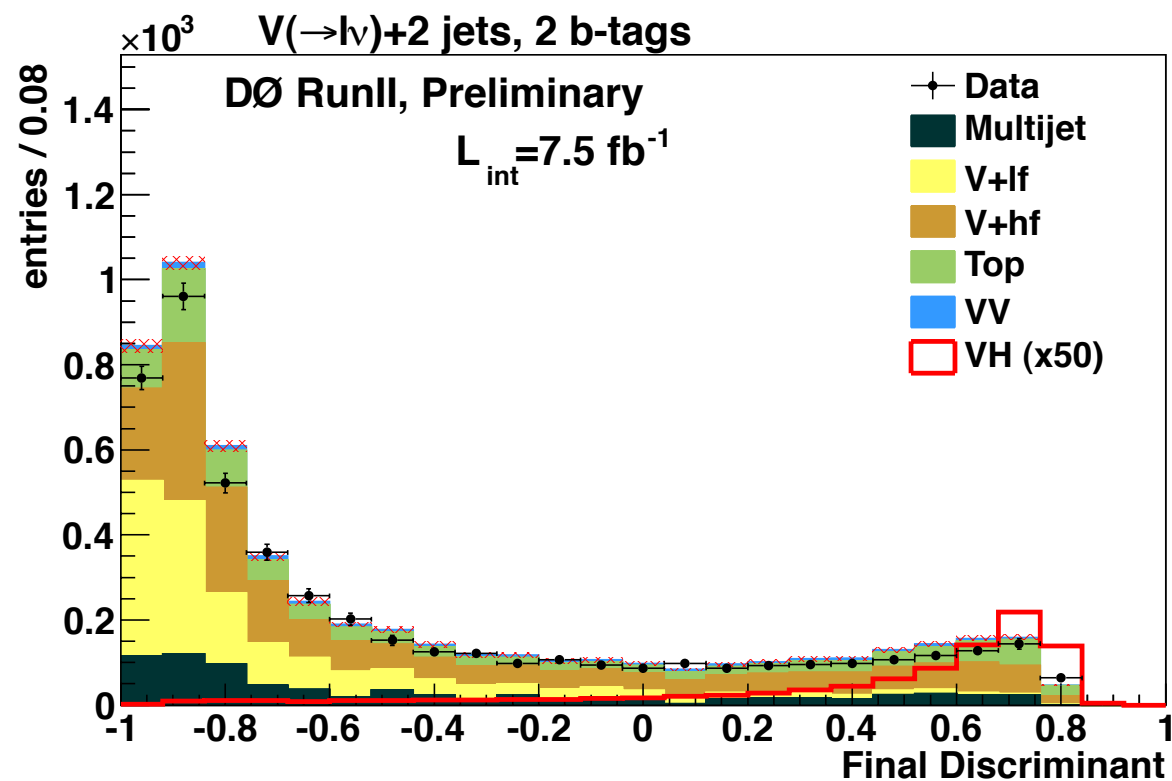
Separate signal and background
before statistical analysis



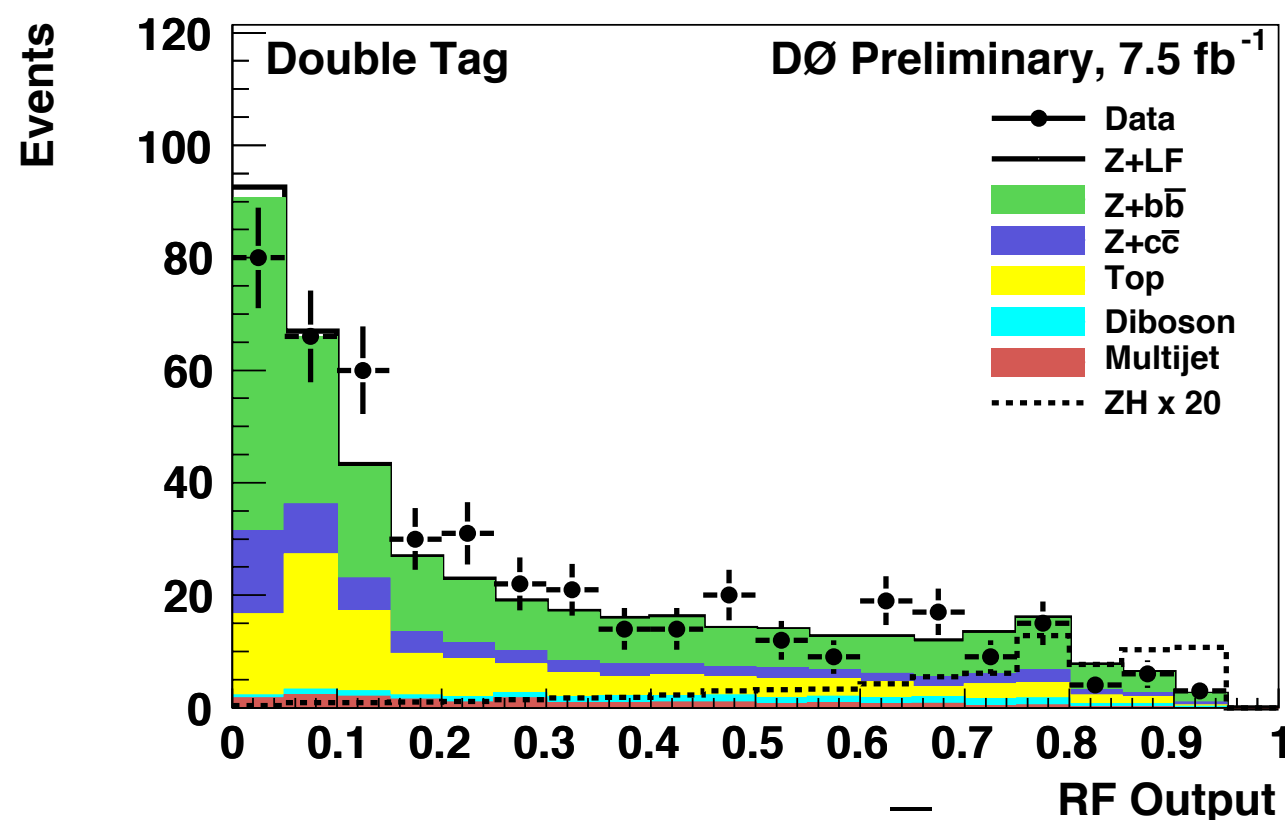
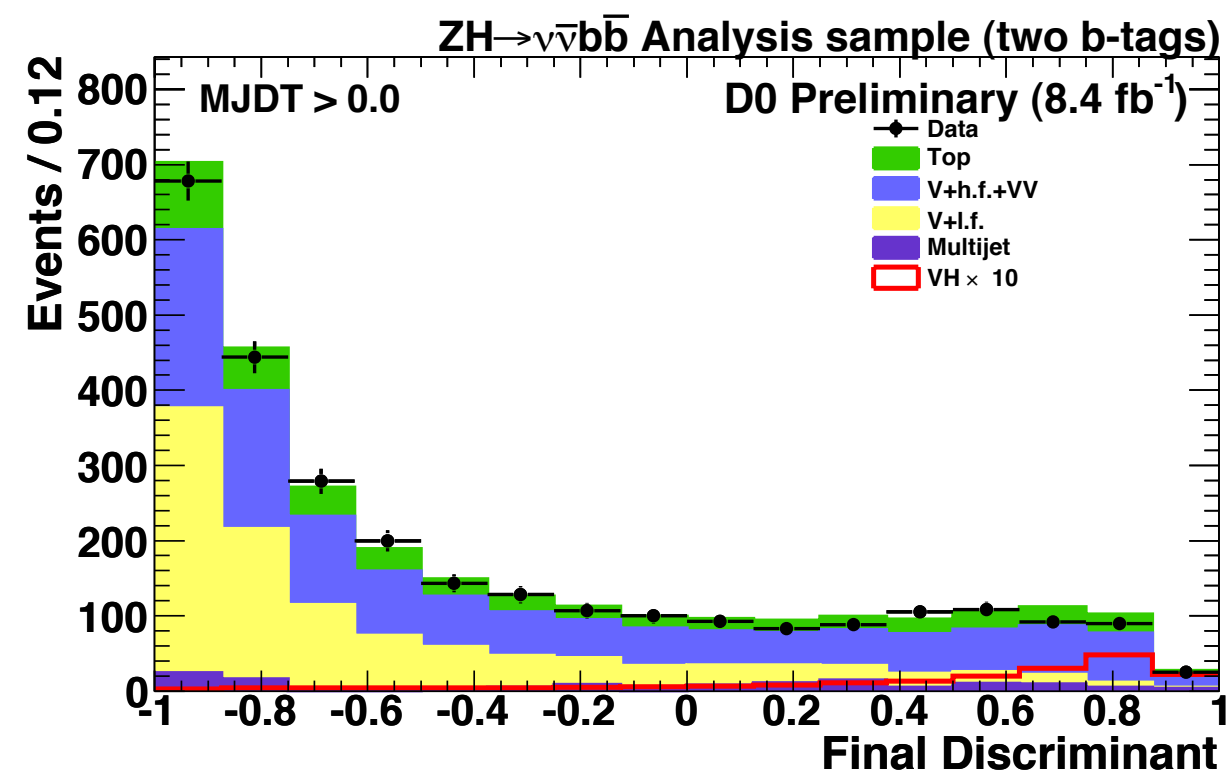
Major LM Inputs ($m_H=115$ GeV)



$$W^\pm H \rightarrow l^\pm \nu b \bar{b}$$

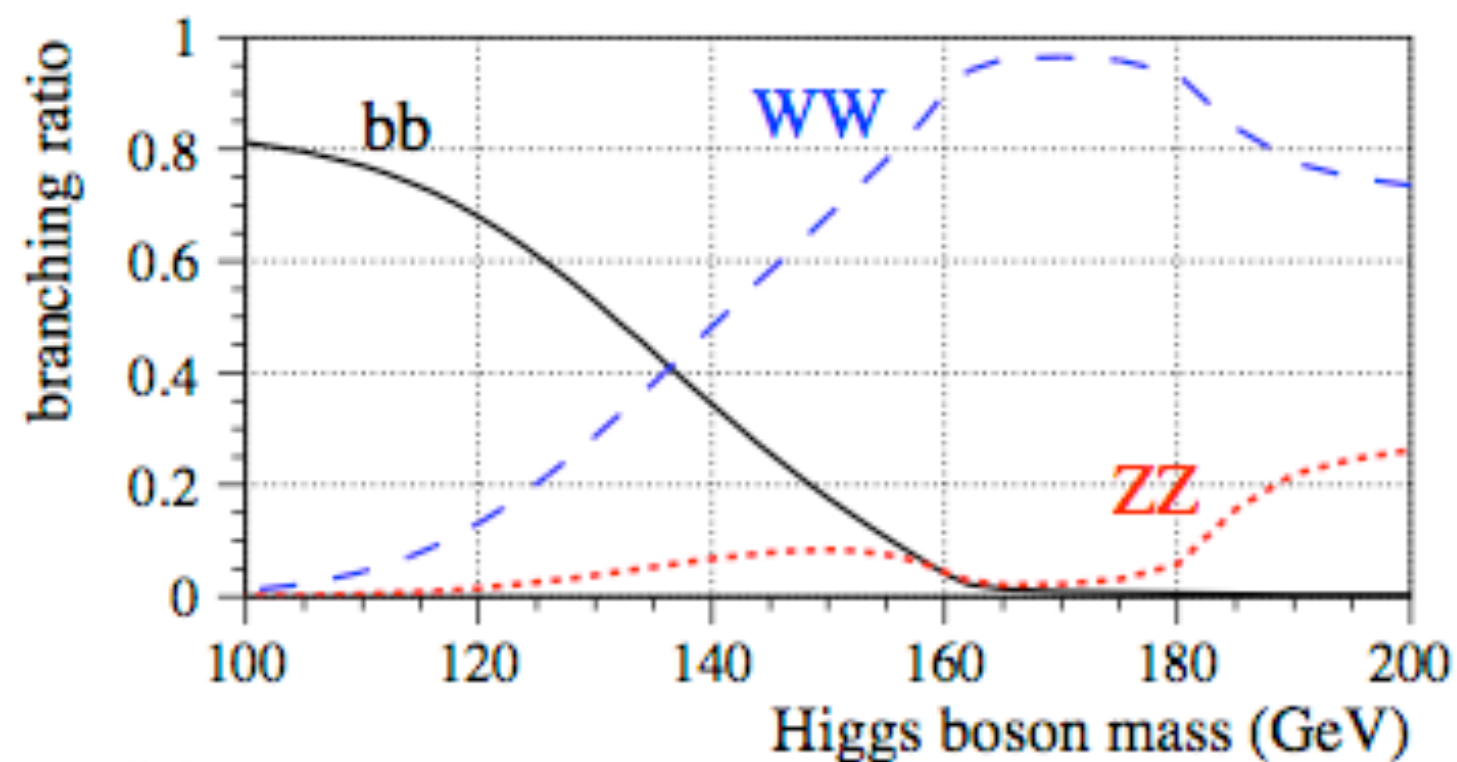


$$ZH \rightarrow \nu \bar{\nu} b \bar{b}$$



$$ZH \rightarrow l^+ l^- b \bar{b}$$

SM branching ratios



high mass : $H \rightarrow W^+W^-$ takes over

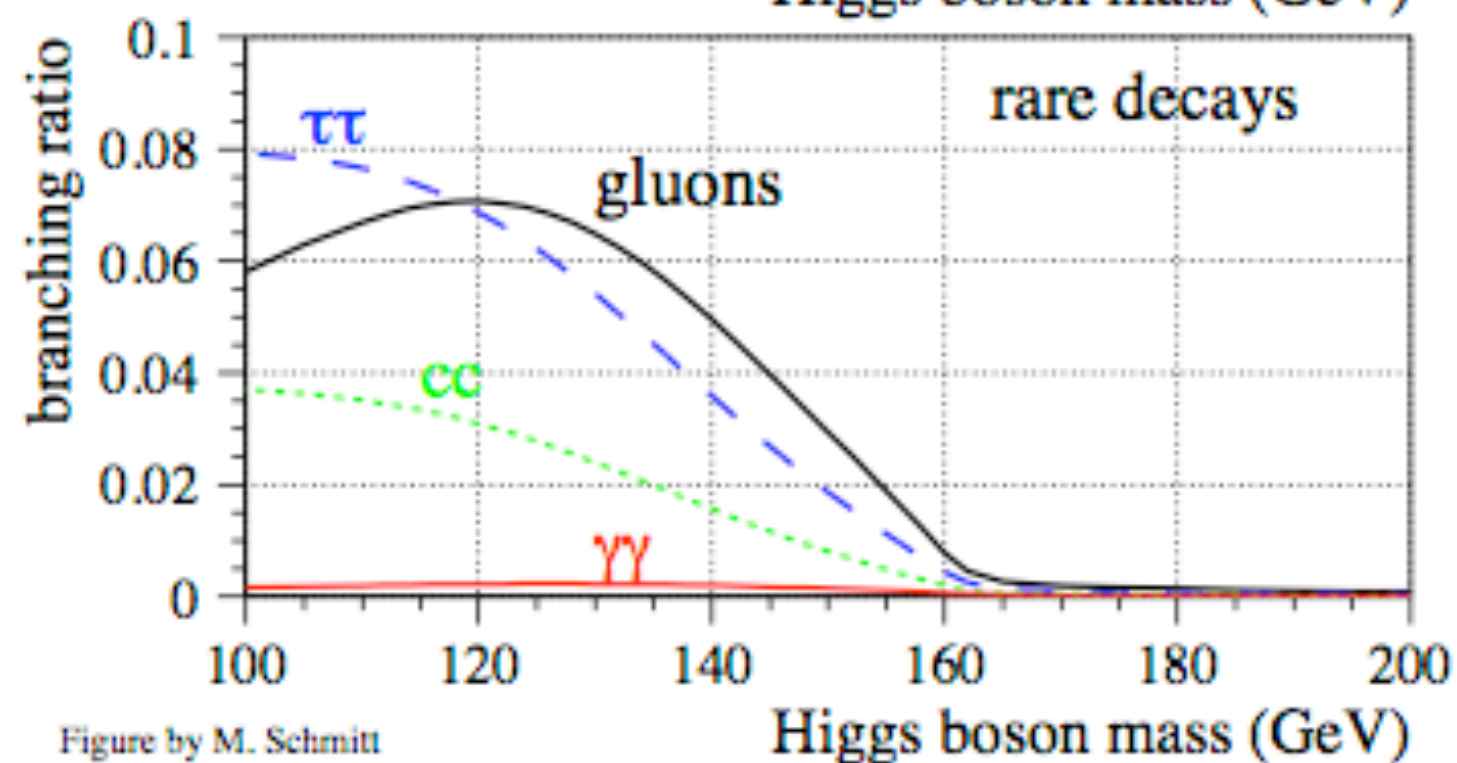


Figure by M. Schmitt

High Mass Signal



For $m_H > 135$ GeV primary decay is $H \rightarrow W^+W^-$

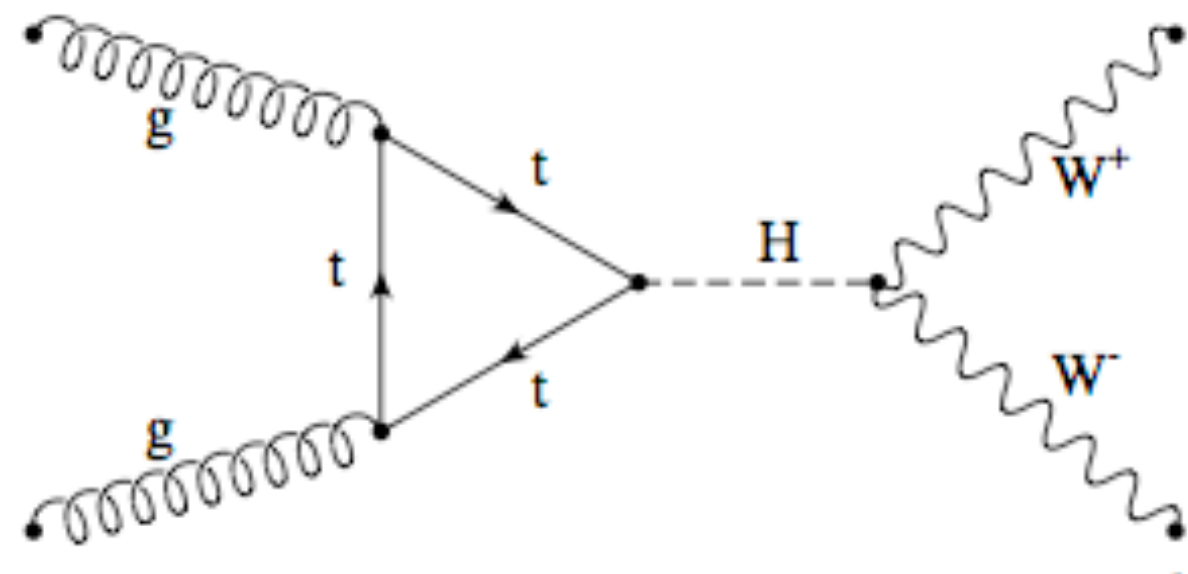
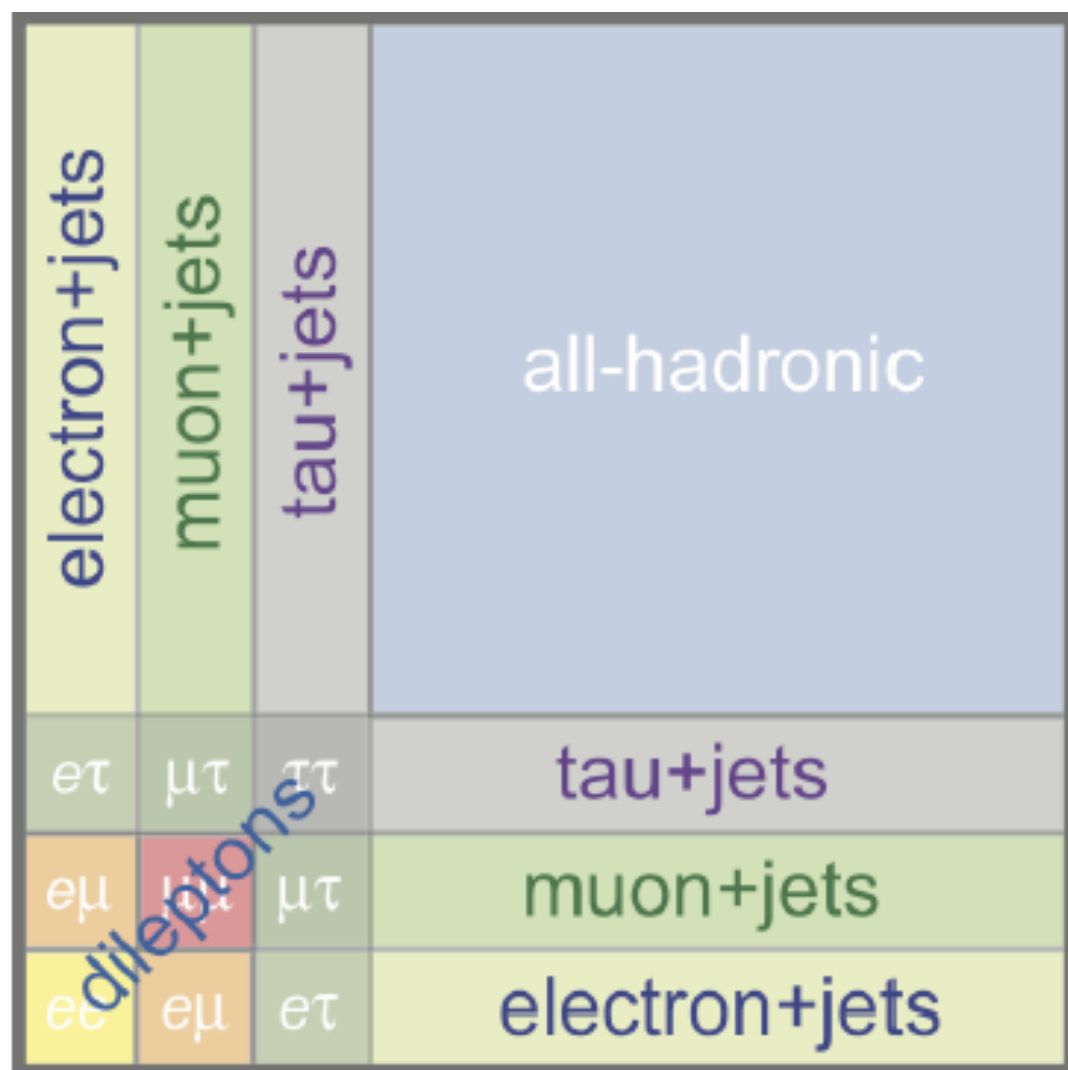
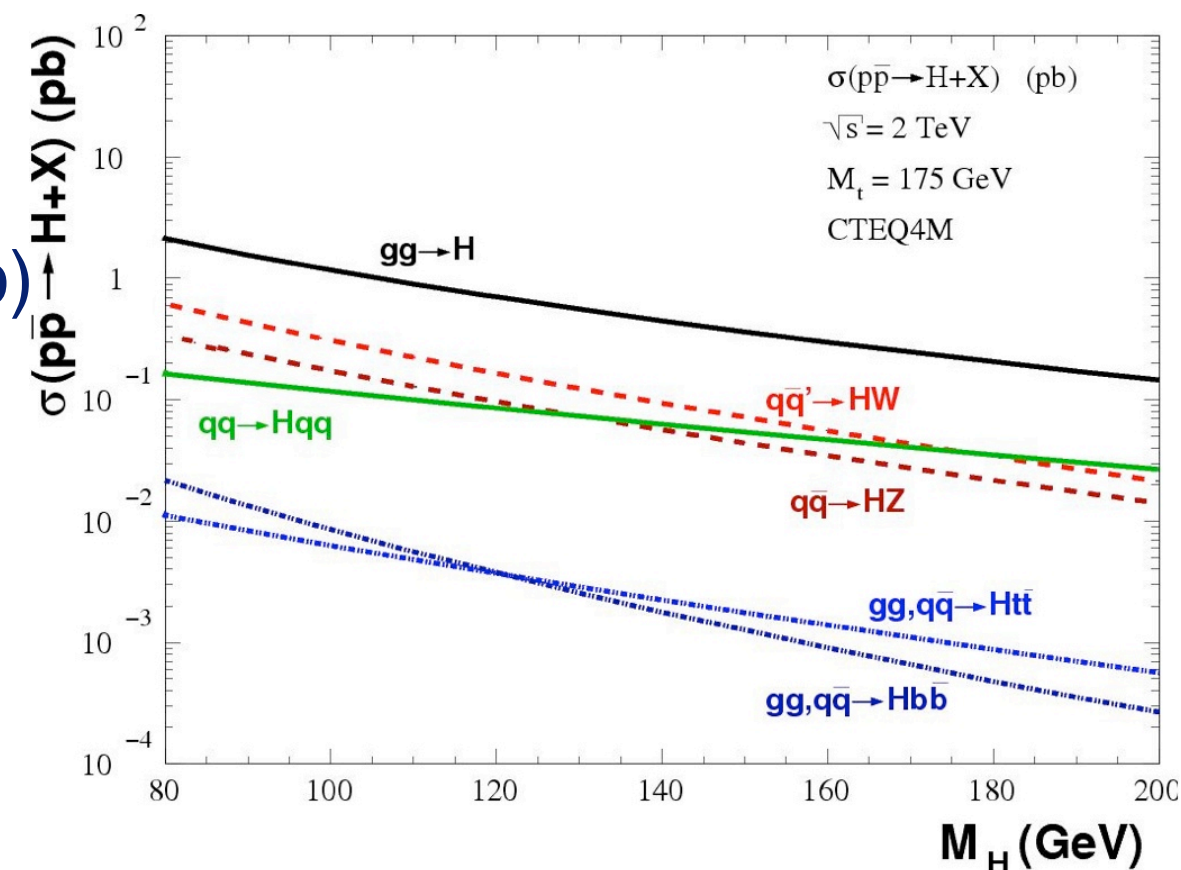
Sensitivity driven $gg \rightarrow H \rightarrow W^+W^- \rightarrow l^+\nu l^-\nu$ (OS lep)

Additional contributions from

$(W/Z)H \rightarrow l^\pm l^\pm + X$ (SS leptons)

$H \rightarrow W^+W^- \rightarrow l^+\nu jj$

and those analysis involving τ decays



High Mass Approach



$H \rightarrow W^+W^- \rightarrow l^+\nu l^-\bar{\nu}$ dominates the sensitivity

Divide samples by lepton flavor and jet multiplicity

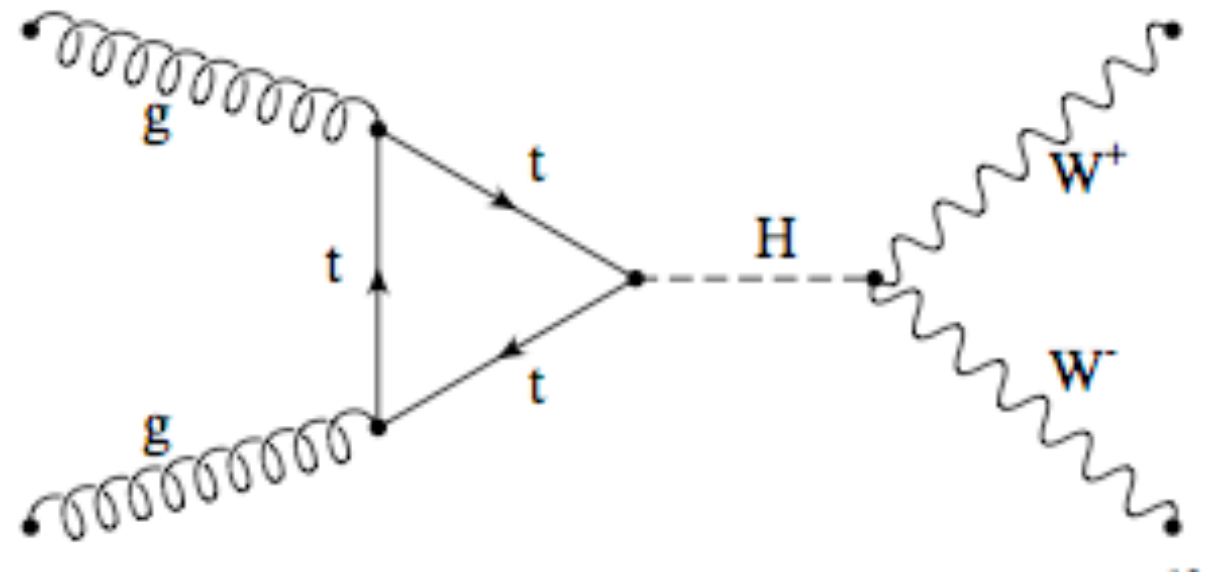
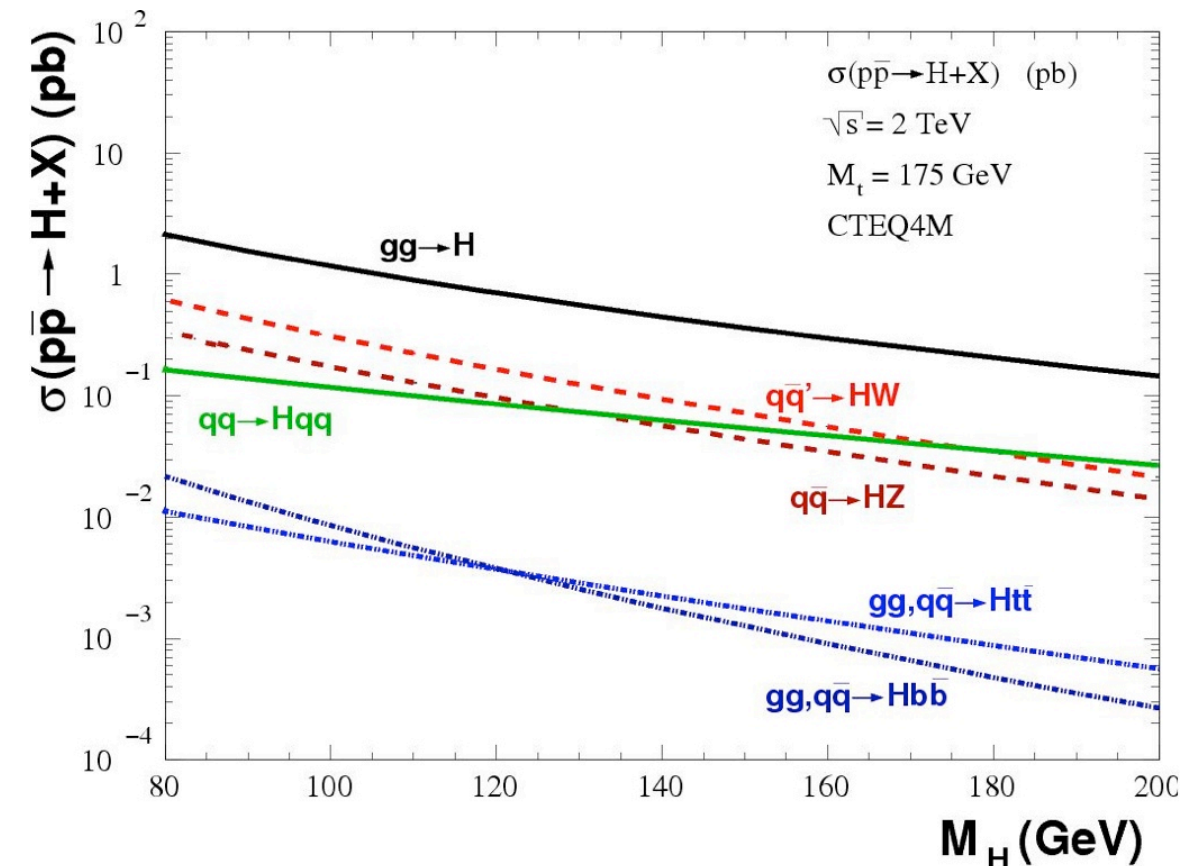
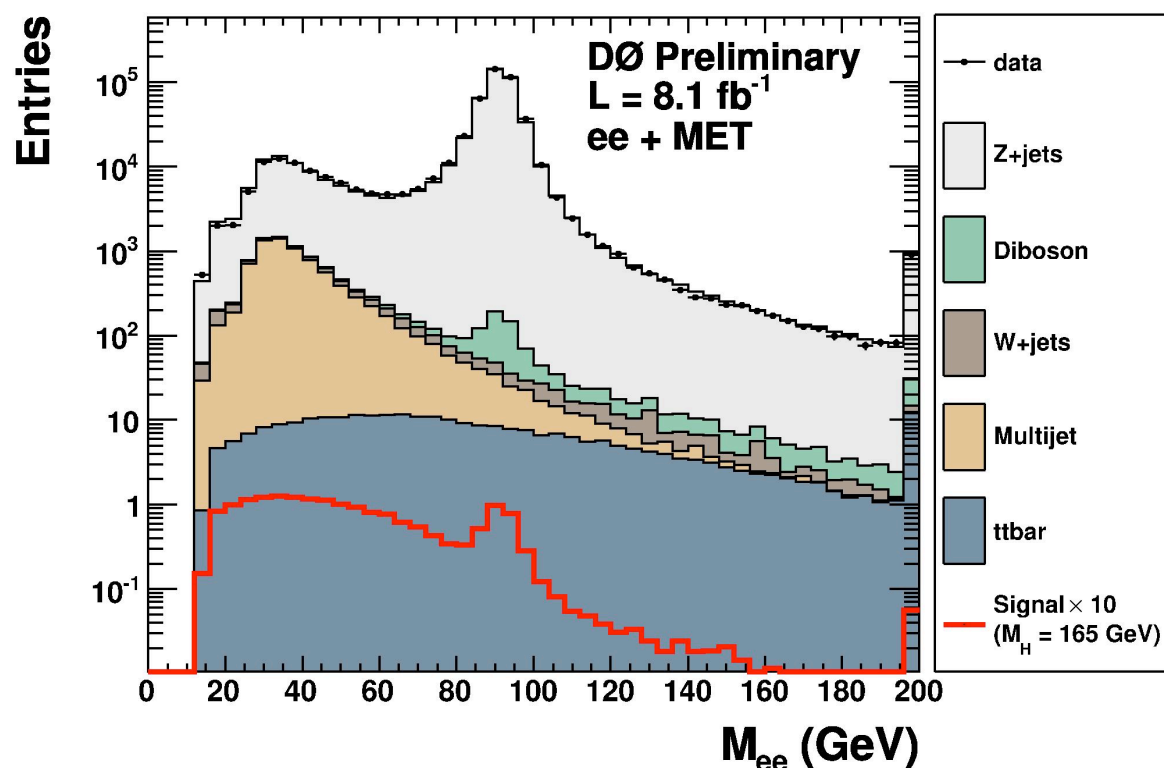
Data driven instrumental background estimations

Multivariate techniques

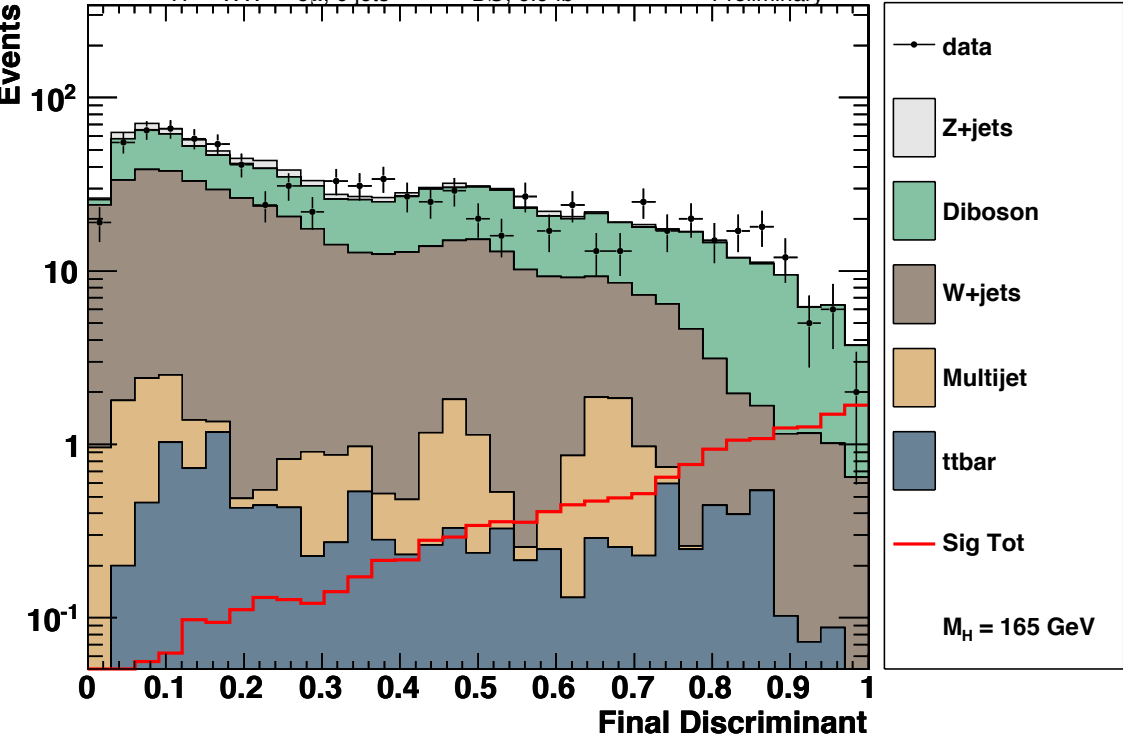
Remove dominant background

Separate signal and background before statistical analysis

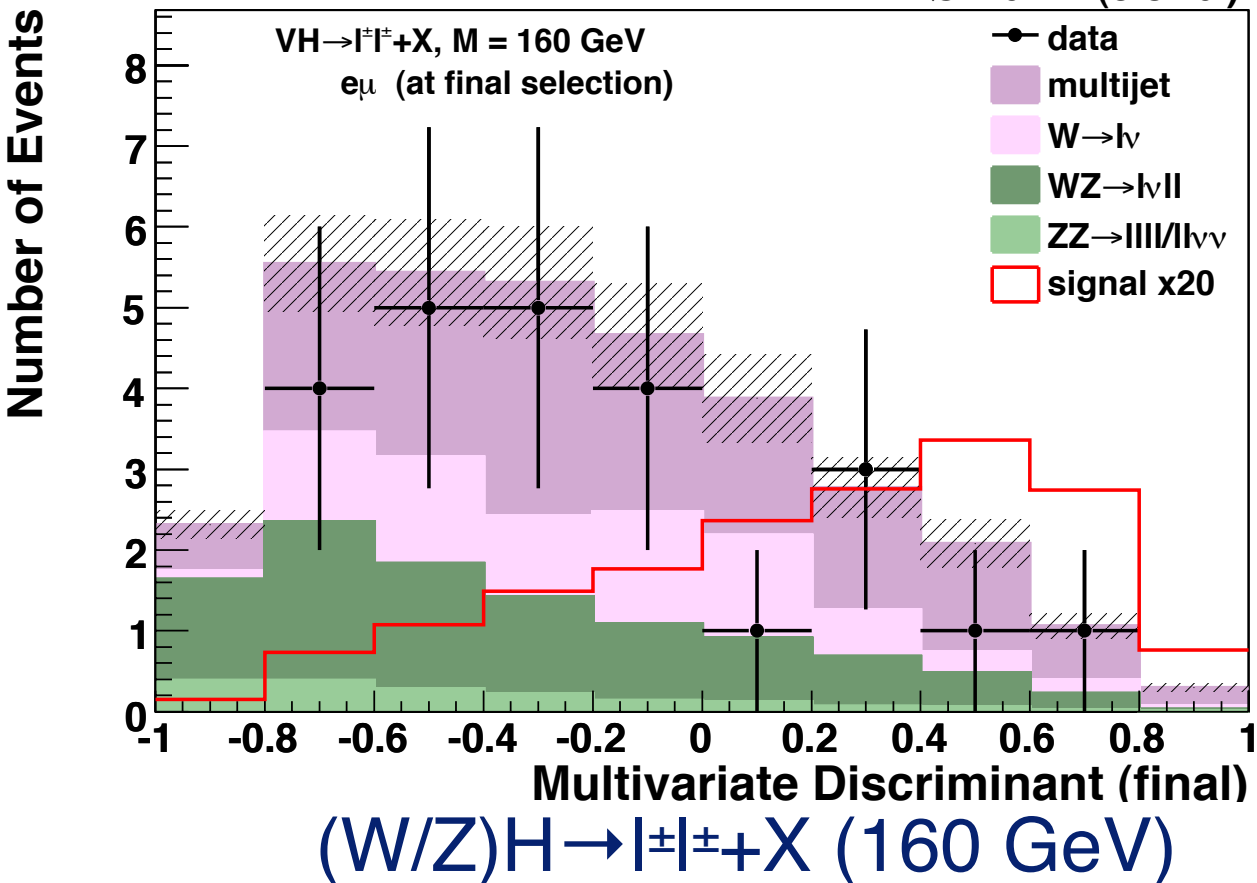
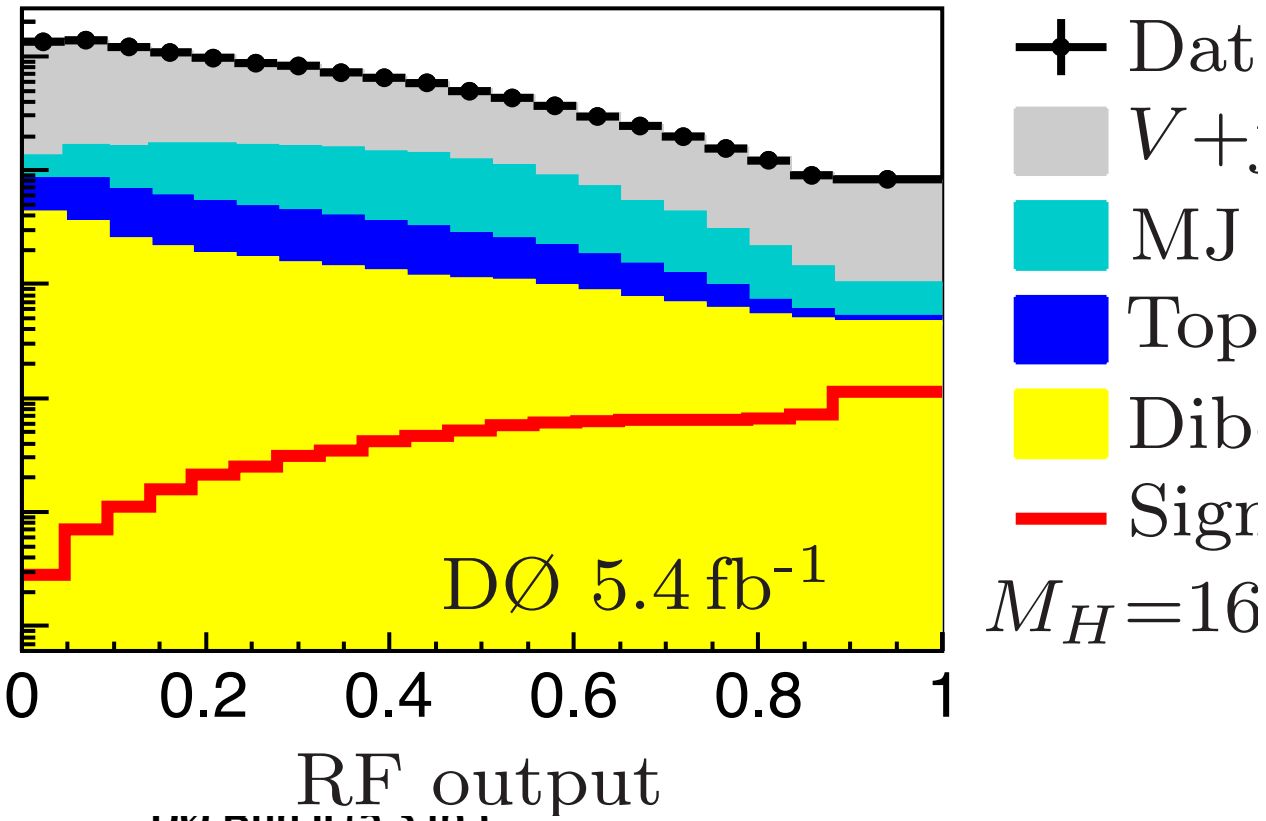
(See Ruchika Nayyar talk)



$H \rightarrow W^+W^- \rightarrow l\nu l\nu$ (165 GeV)



$H \rightarrow W^+W^- \rightarrow l\nu qq$ (165 GeV)



Two SM Higgs searches use *hadronically* decaying taus

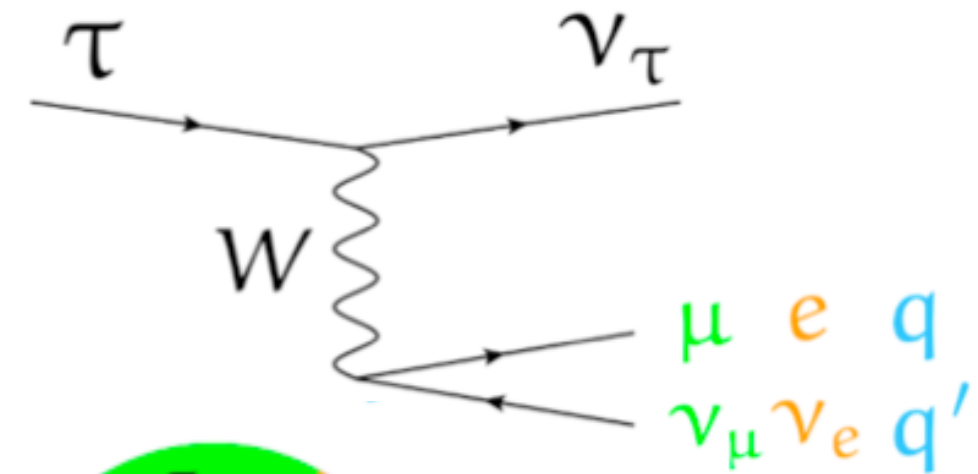
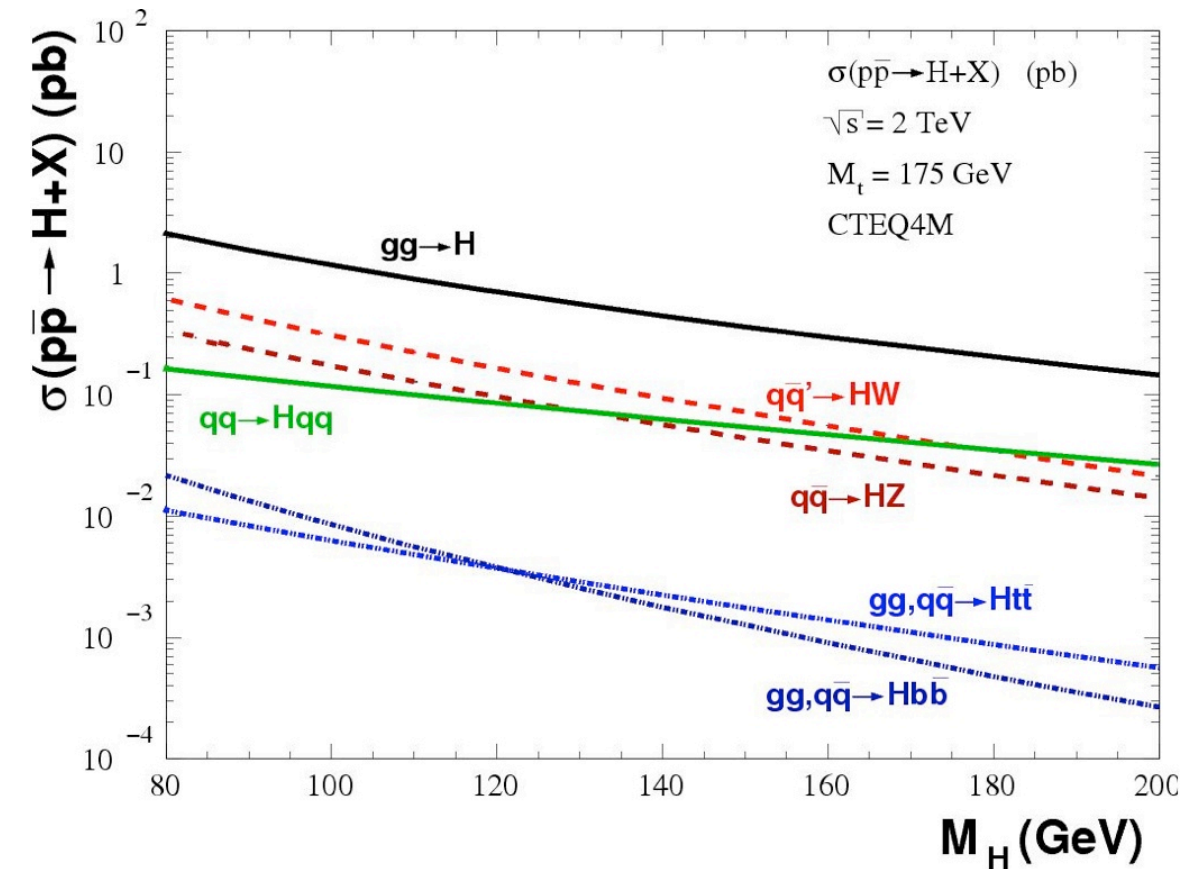
$$H \rightarrow W+W^- \rightarrow \mu\nu\tau\nu$$

$$H+X \rightarrow (e/\mu)\tau+qq$$

Leptonically decaying taus used in lepton channels.

Low mass - analyze neglected production methods i.e. $gg \rightarrow H$

High mass - Increase signal acceptance



BR_τ



Object Selection

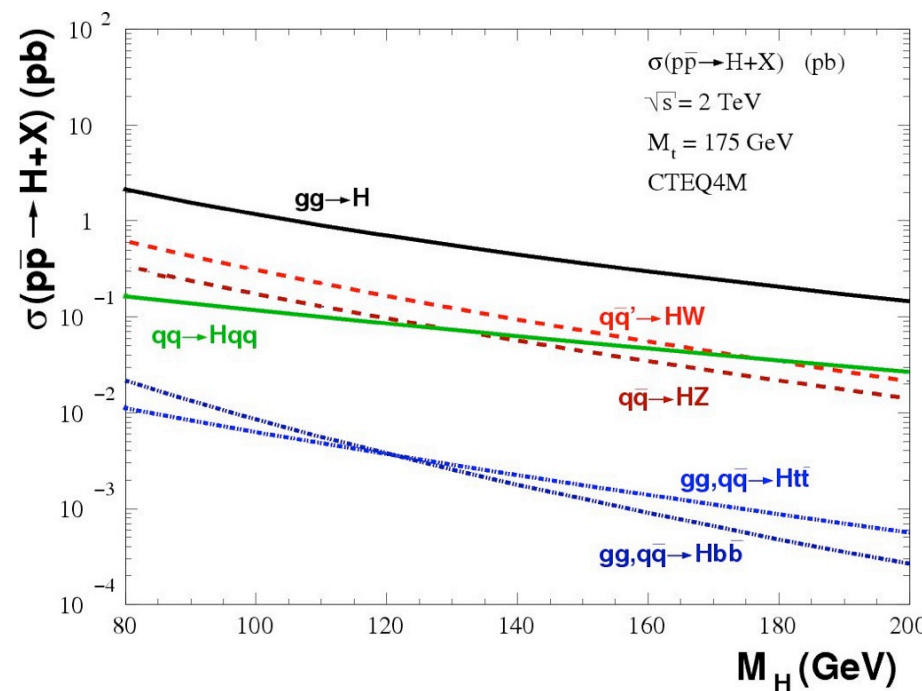
- 1 high p_T isolated lepton
- 1 hadronically decaying tau
- 2 jets

Control Samples

- W +jets - Alter tau selection
- Z +jets - Replace tau with muon

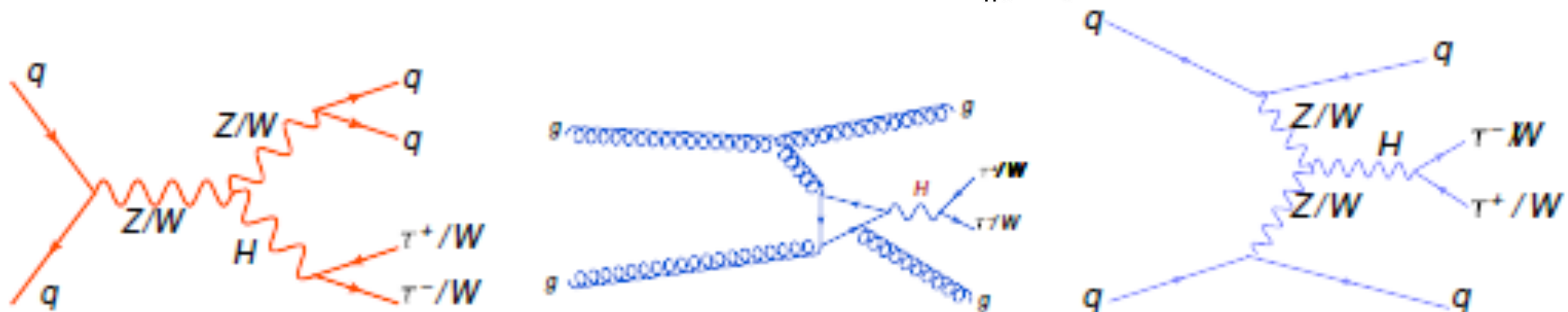
Low mass signals

- $gg \rightarrow H \rightarrow \tau\tau + jj$
- $qq \rightarrow (W \rightarrow jj)(H \rightarrow \tau\tau)$
- $qq \rightarrow (Z \rightarrow jj)(H \rightarrow \tau\tau)$
- $qq' \rightarrow qq'(H \rightarrow \tau\tau)$
- $qq' \rightarrow (B \rightarrow bb)(Z \rightarrow \tau\tau)$



High mass signals

- $qq \rightarrow W(H \rightarrow W+W-)$
- $gg \rightarrow H \rightarrow W+W- + jj$
- $qq \rightarrow Z(H \rightarrow W+W-)$
- $qq' \rightarrow qq'(H \rightarrow W+W-)$

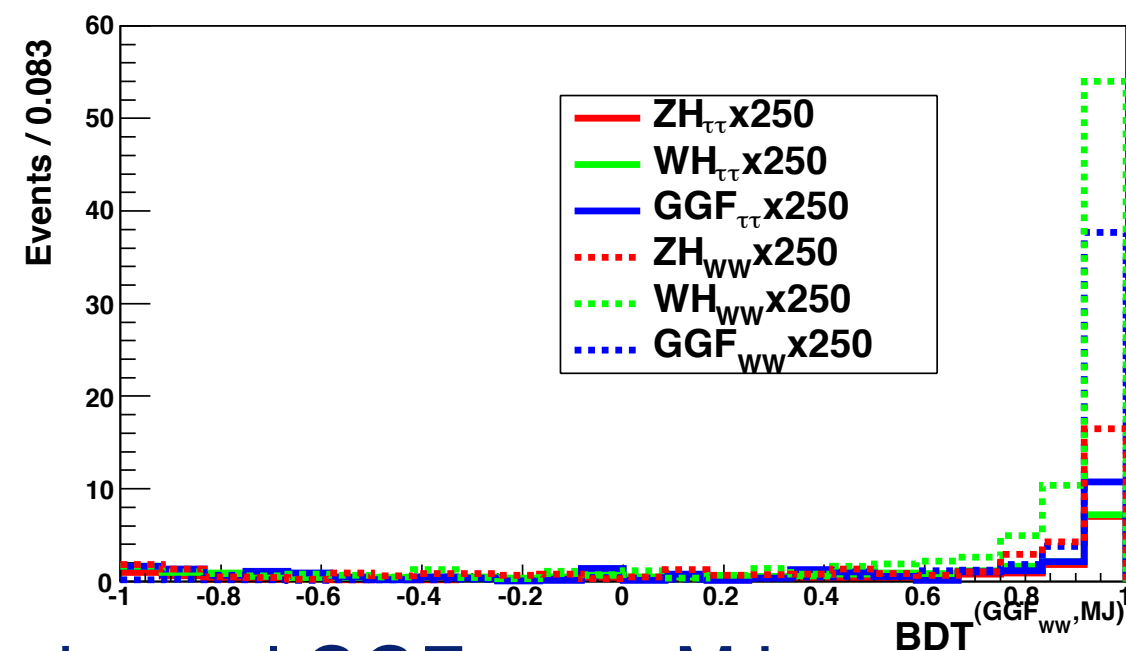
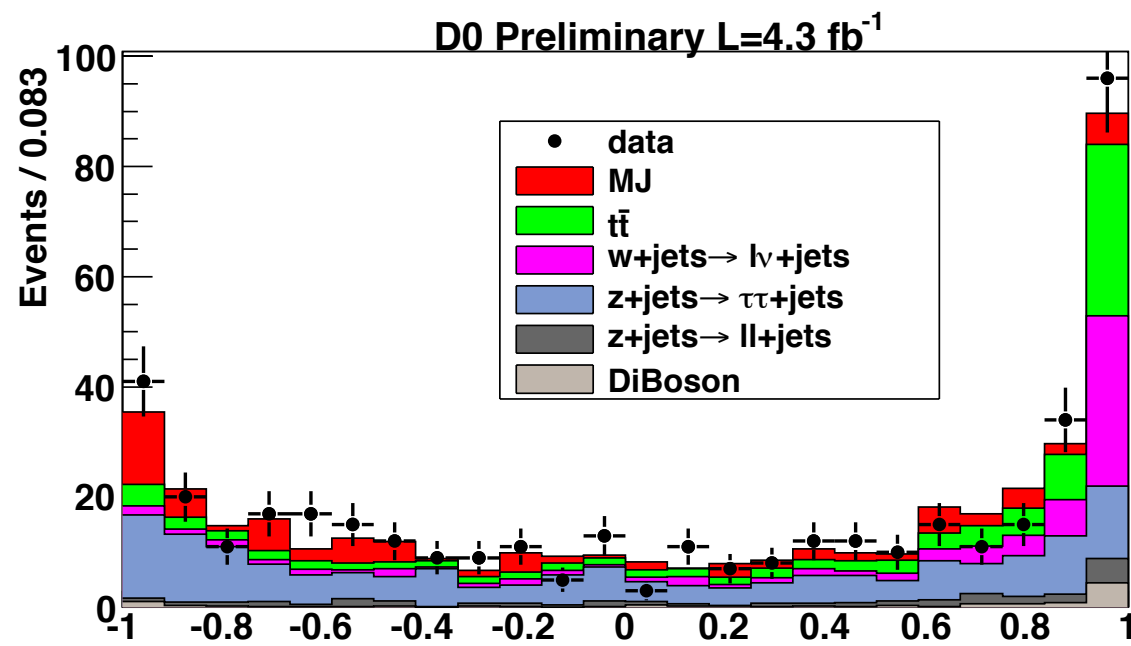


The lepton can be produced by $Z \rightarrow \ell\ell$, $W \rightarrow \ell\nu$, $\tau \rightarrow \ell\nu\nu$

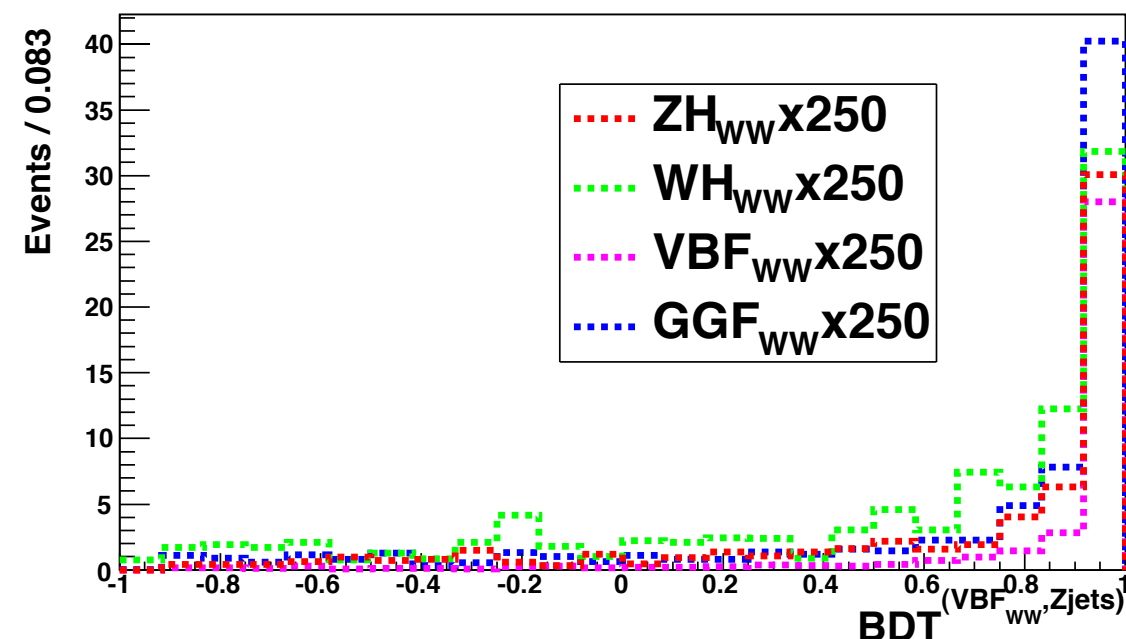
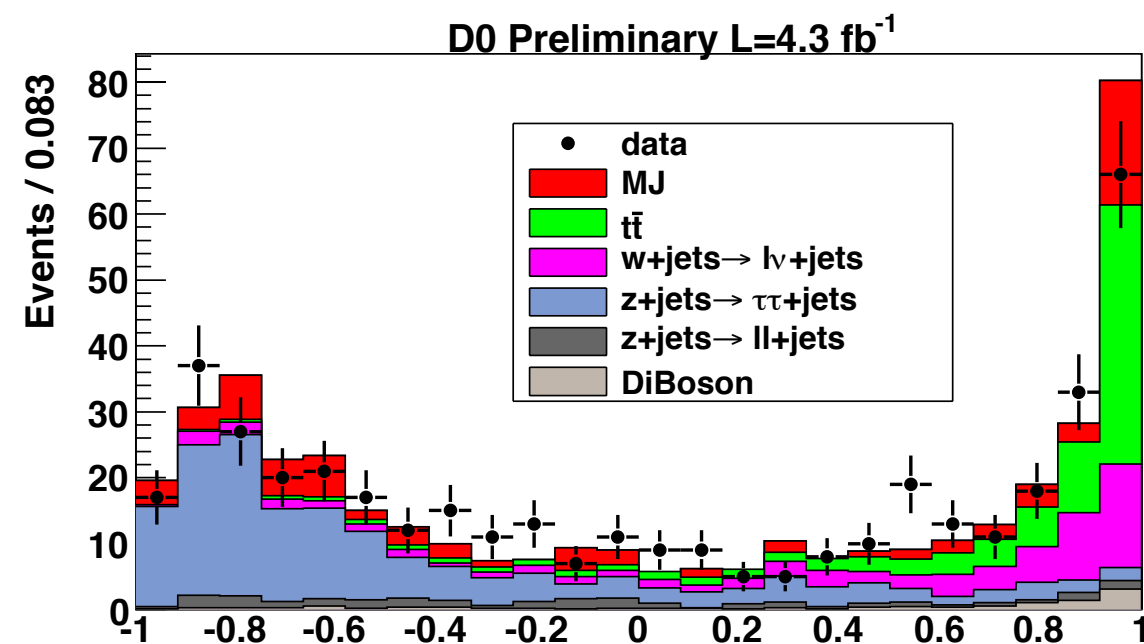
$H+X \rightarrow (e/\mu)\tau+qq$ Decision Trees



3 mass regions $m_H < 125$, $125 \leq m_H \leq 135$, $m_H > 135$ GeV - signal composition differs
Train a BDT for each of the major signals against a major background

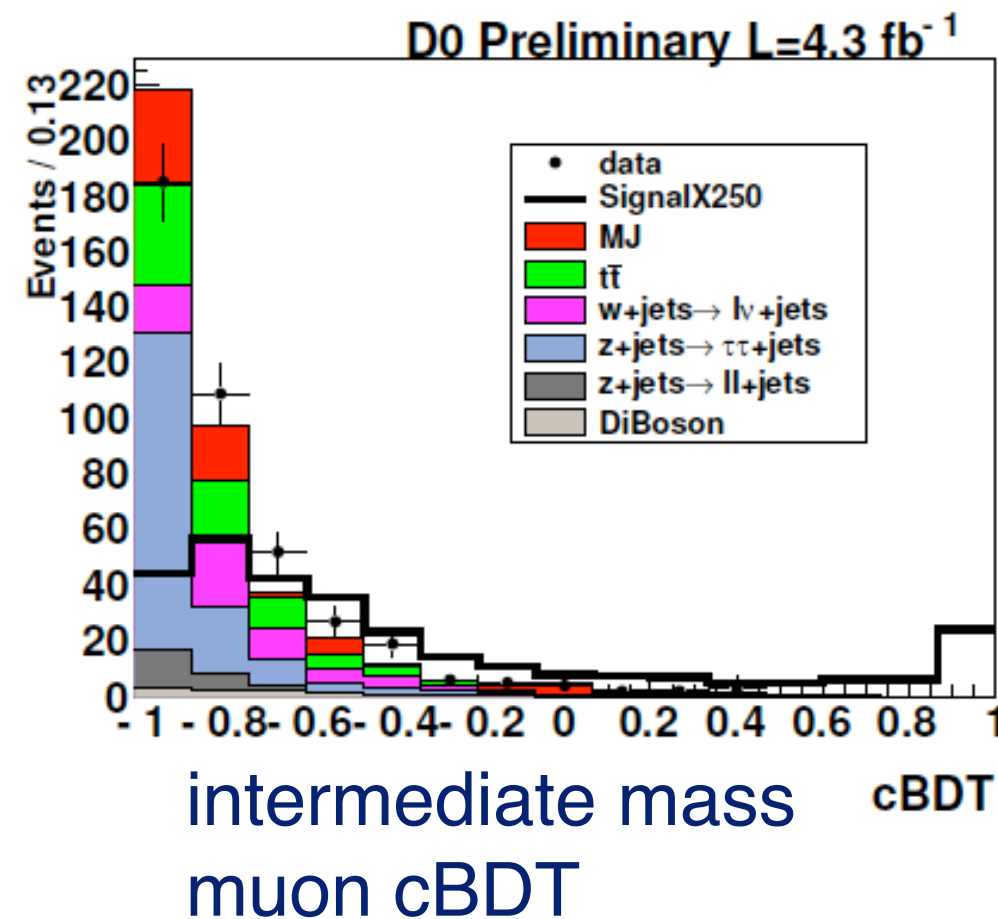
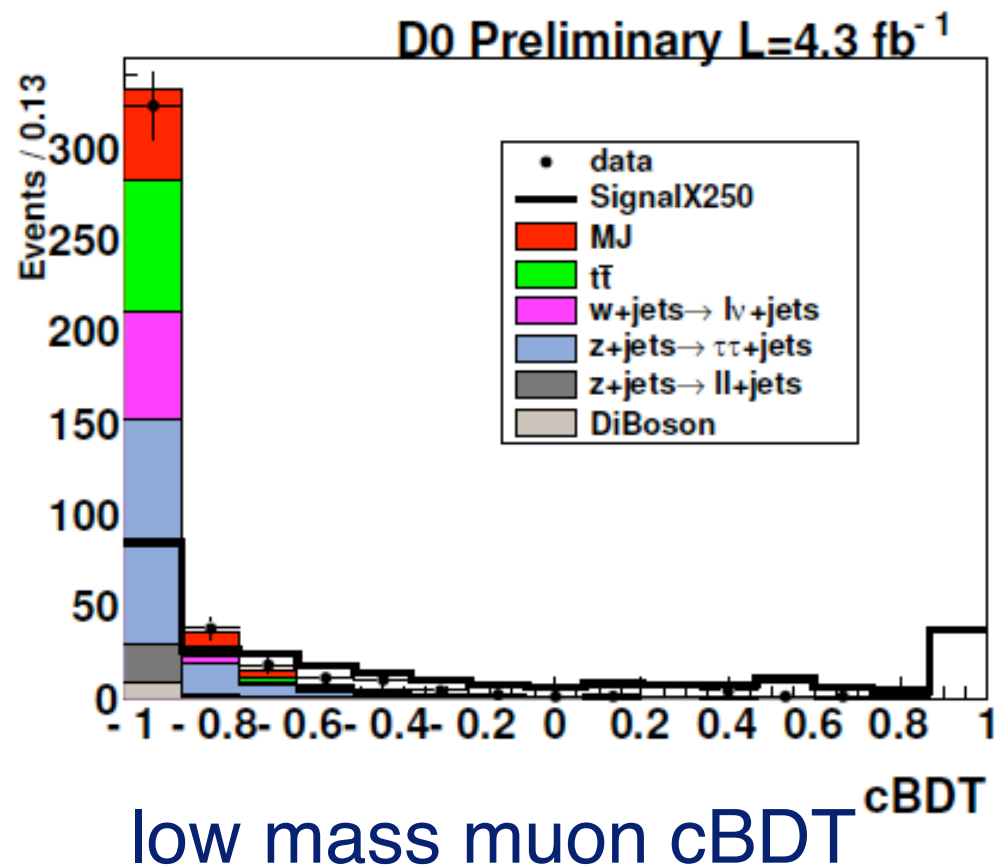


intermediate mass muon channel GGF_{WW} vs MJ



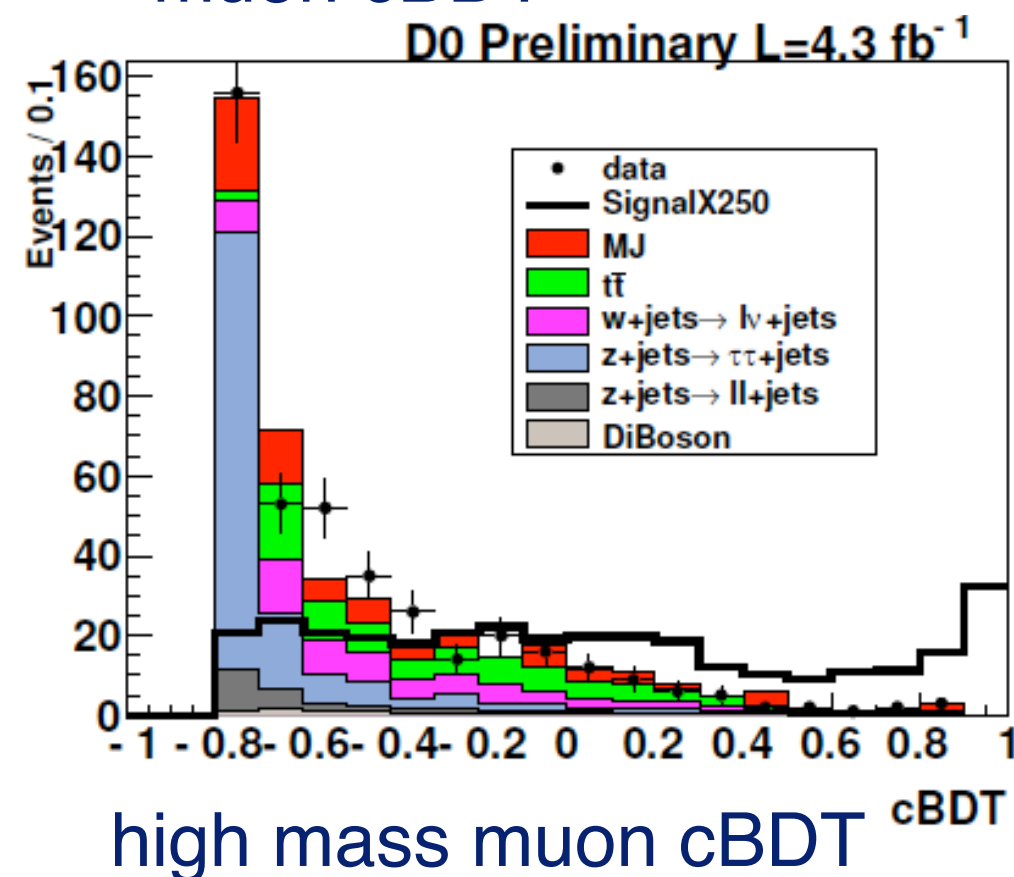
high mass muon channel VBF_{WW} vs Zjet

$H+X \rightarrow (e/\mu)\tau+qq$ Decision Trees



Give individual DTs as an inputs to a combined BDT (cBDT)

Higgs mass region	Signals				
low	$GGF_{\tau\tau}$	$VH_{\tau\tau}$	$VBF_{\tau\tau}$		
intermediate	$GGF_{\tau\tau}$	GGF_{WW}	$VH_{\tau\tau}$	VH_{WW}	
high	GGF_{WW}	VH_{WW}	VBF_{WW}		



How are Higgs Production Limits Set?



A modified frequentist approach is used.
The CL_s method.

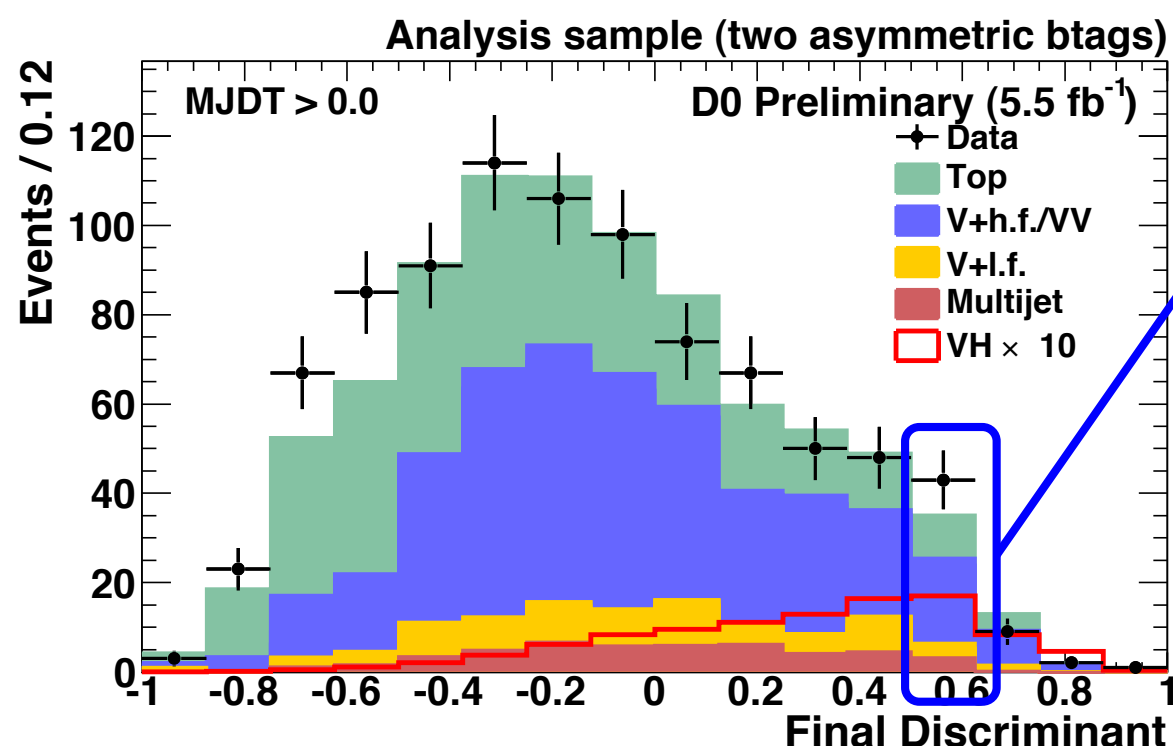
Compare Poisson likelihood of Background (B) hypothesis to Signal+Background (S+B) hypothesis and calculate their **negative log likelihood ratio (LLR)**

$L(B)$	$L(S + B)$	LLR
$\prod_i \frac{b_i^{d_i} \exp(-b_i)}{d_i!}$	$\prod_i \frac{(s_i + b_i)^{d_i} \exp(-(s_i + b_i))}{d_i!}$	$2 \cdot \sum_i s_i - d_i \cdot \log(1 + s_i/b_i)$

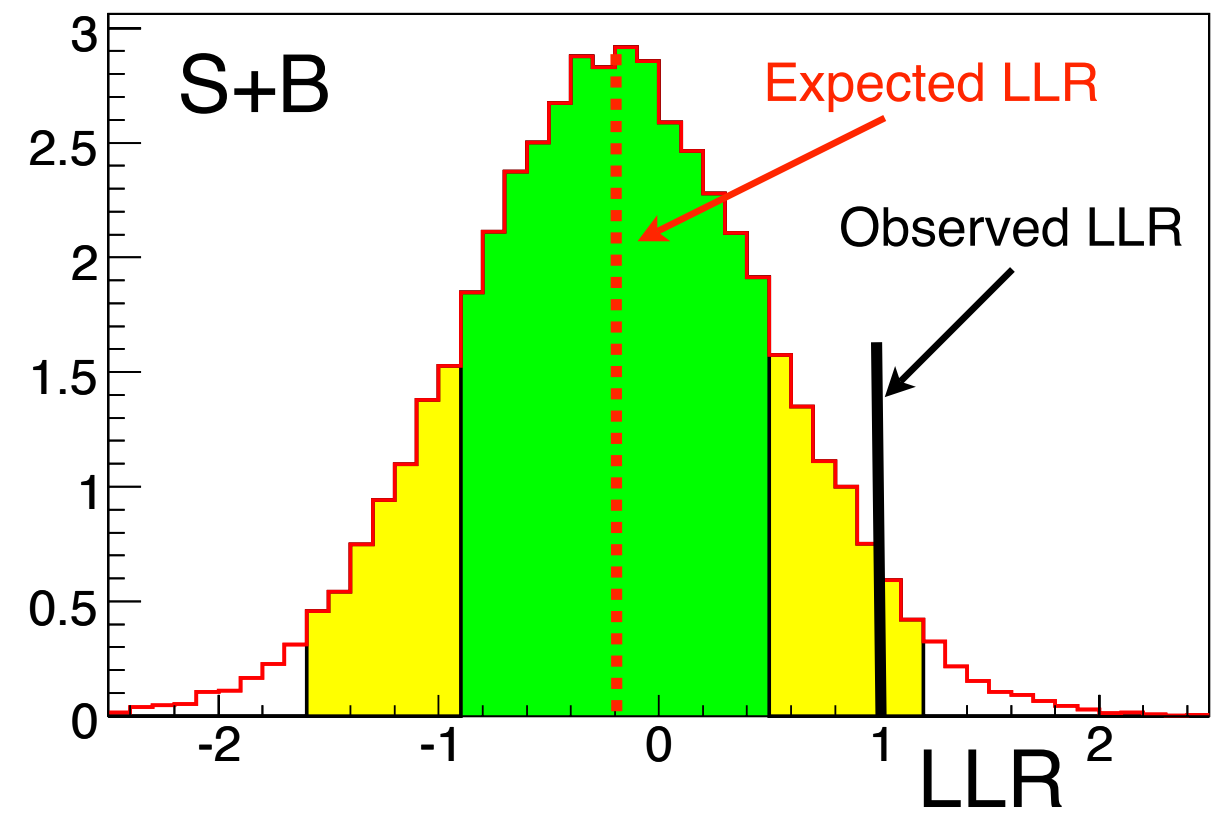
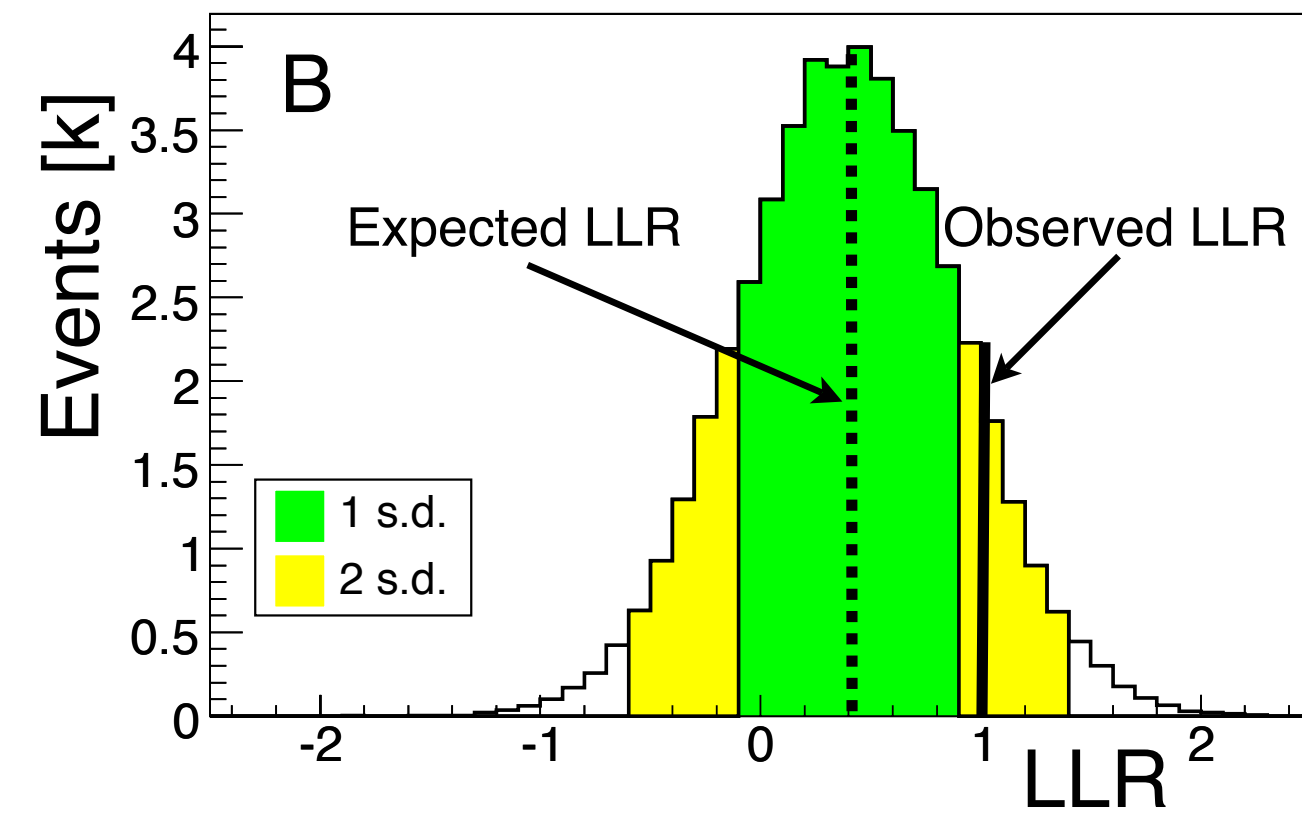
Compare Poisson likelihood of Background (B) hypothesis to Signal+Background (S+B) hypothesis and calculate their **negative log likelihood ratio (LLR)**

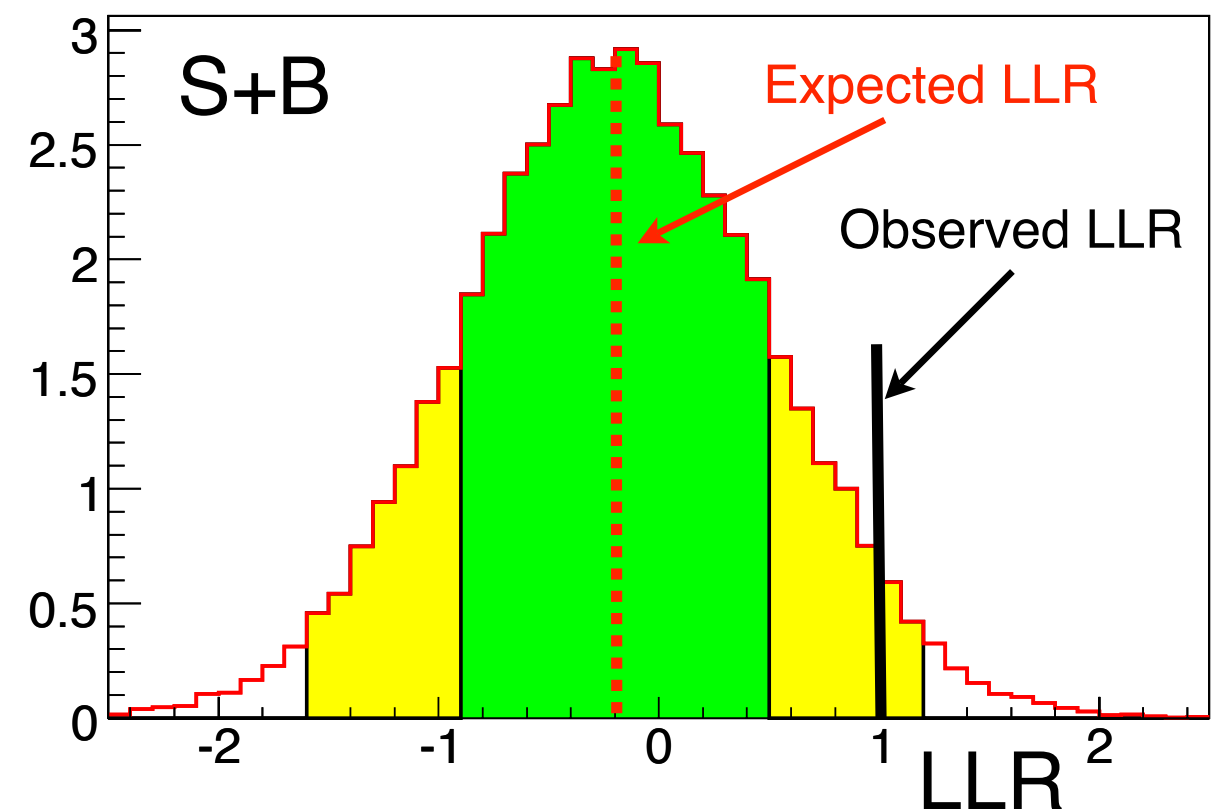
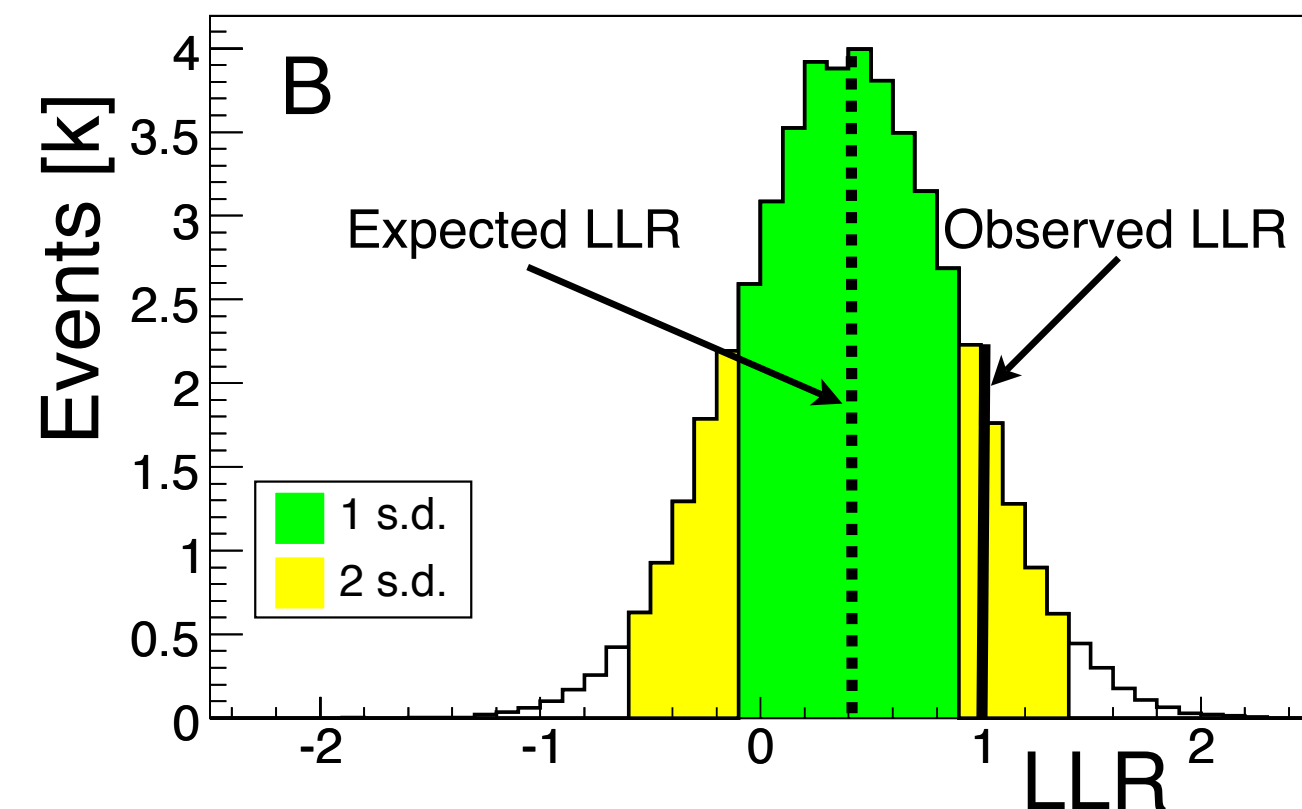
$L(B)$	$L(S + B)$	LLR
$\prod_i \frac{b_i^{d_i} \exp(-b_i)}{d_i!}$	$\prod_i \frac{(s_i + b_i)^{d_i} \exp(-(s_i + b_i))}{d_i!}$	$2 \cdot \sum_i s_i - d_i \cdot \log(1 + s_i/b_i)$

where d_i events observed in bin i with S and B expectations s_i and b_i .



- $d_i = 42, b_i = 35, s_i = 1.5$.
- LLR = -0.52
- Sum overall all bins to compute **Observed LLR**.
- Add additional channels as new bins.

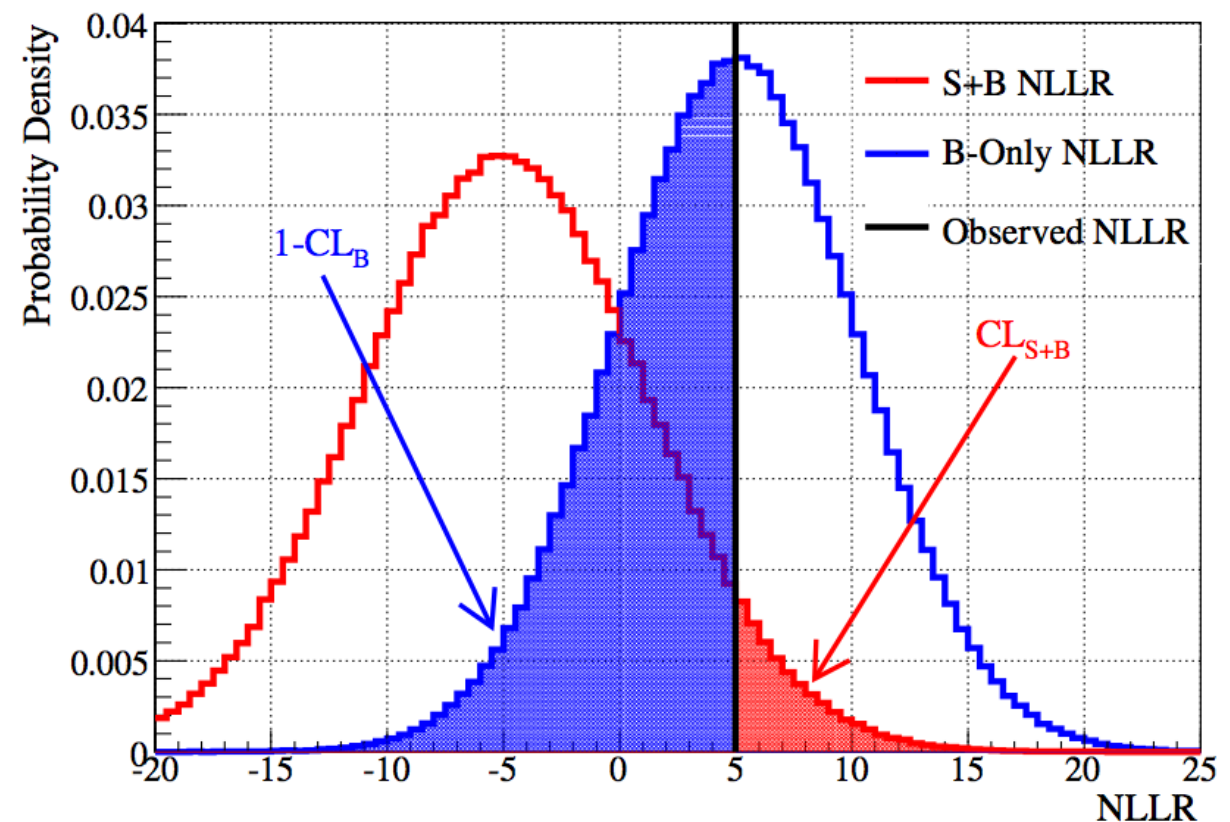




Fold in systematics via
Bayesian marginalization

$CL_s = CL_{sb}/CL_b$ is safe against
highly unlikely B-like fluctuate

CLs over-covers CL_b (95% CL
with $CL_s \sim 98\%$ CL with CL_b)

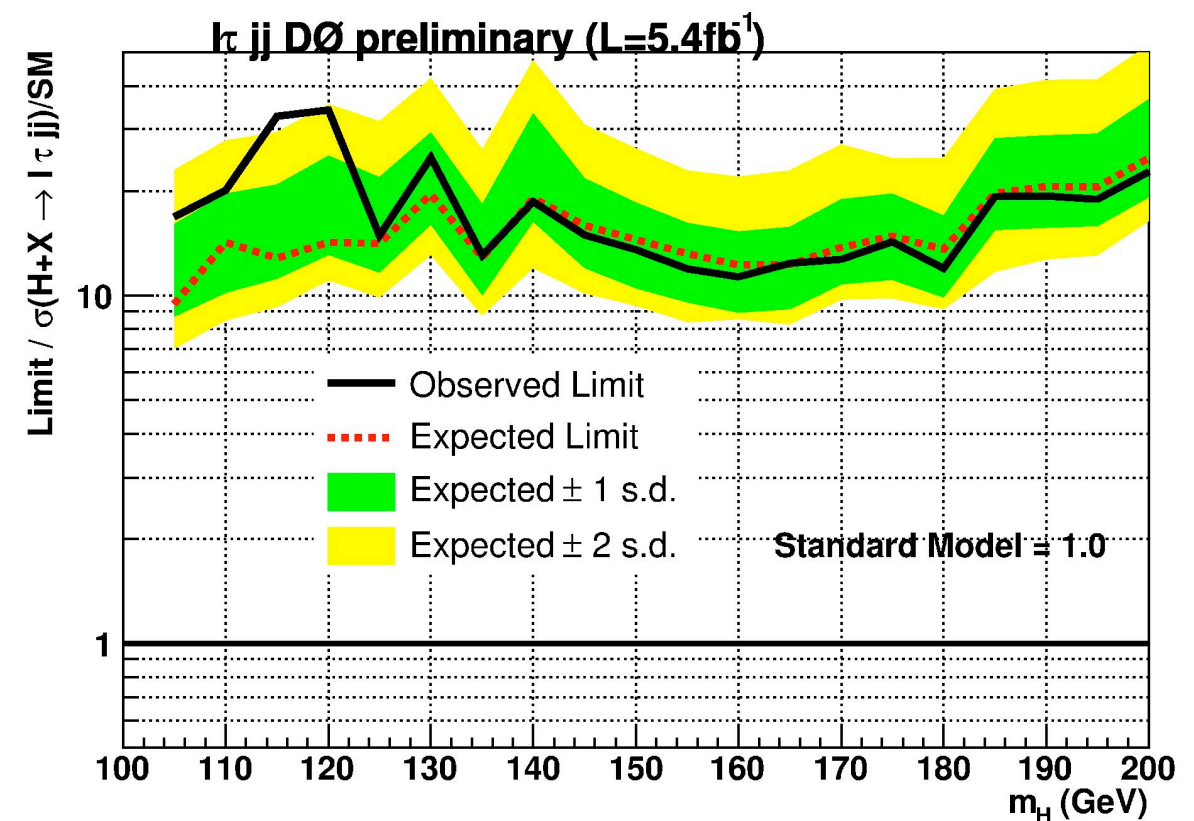
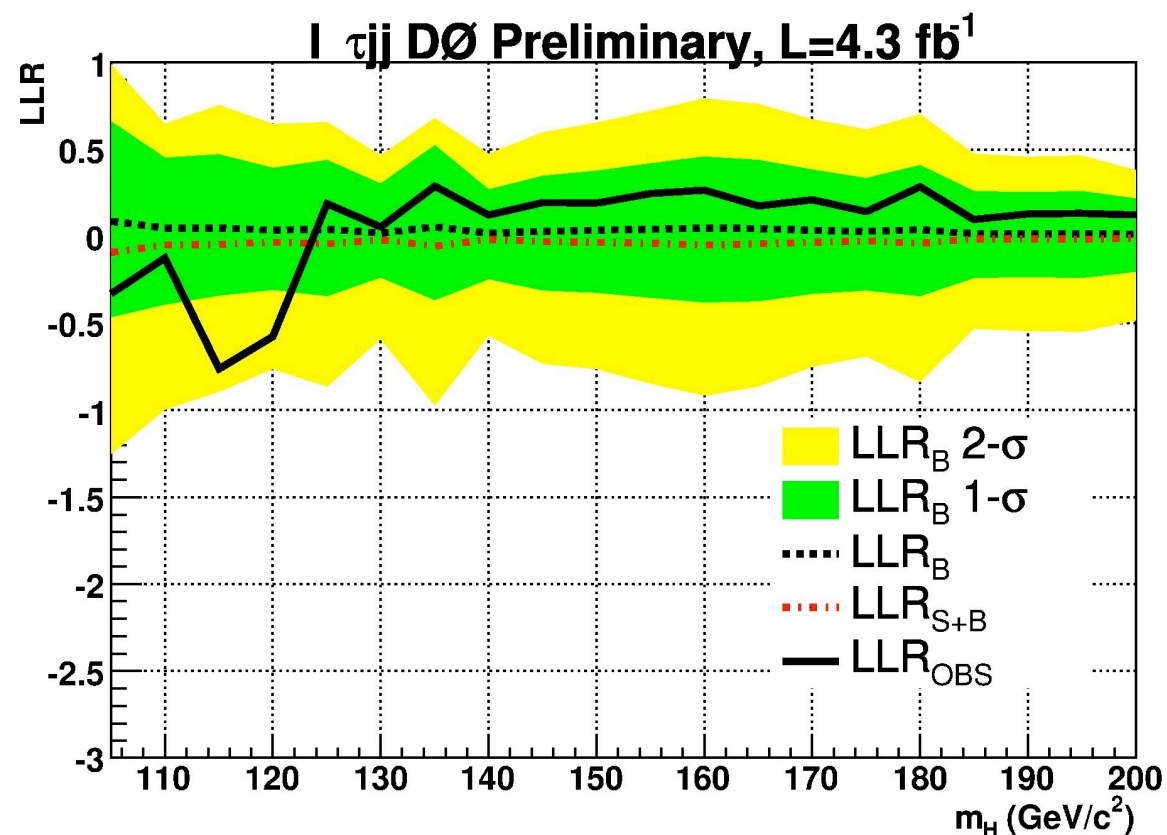


Limits $X+H \rightarrow \tau\tau+jj$

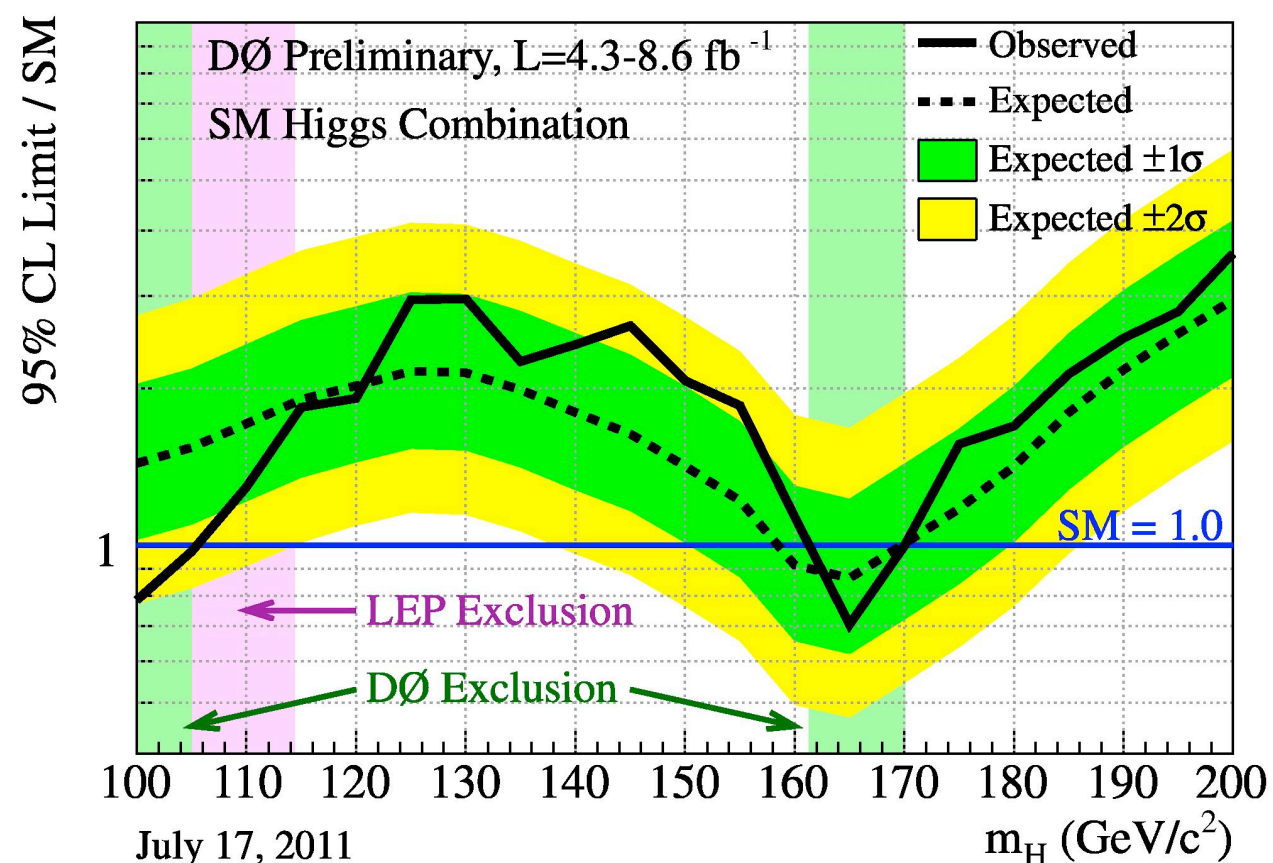
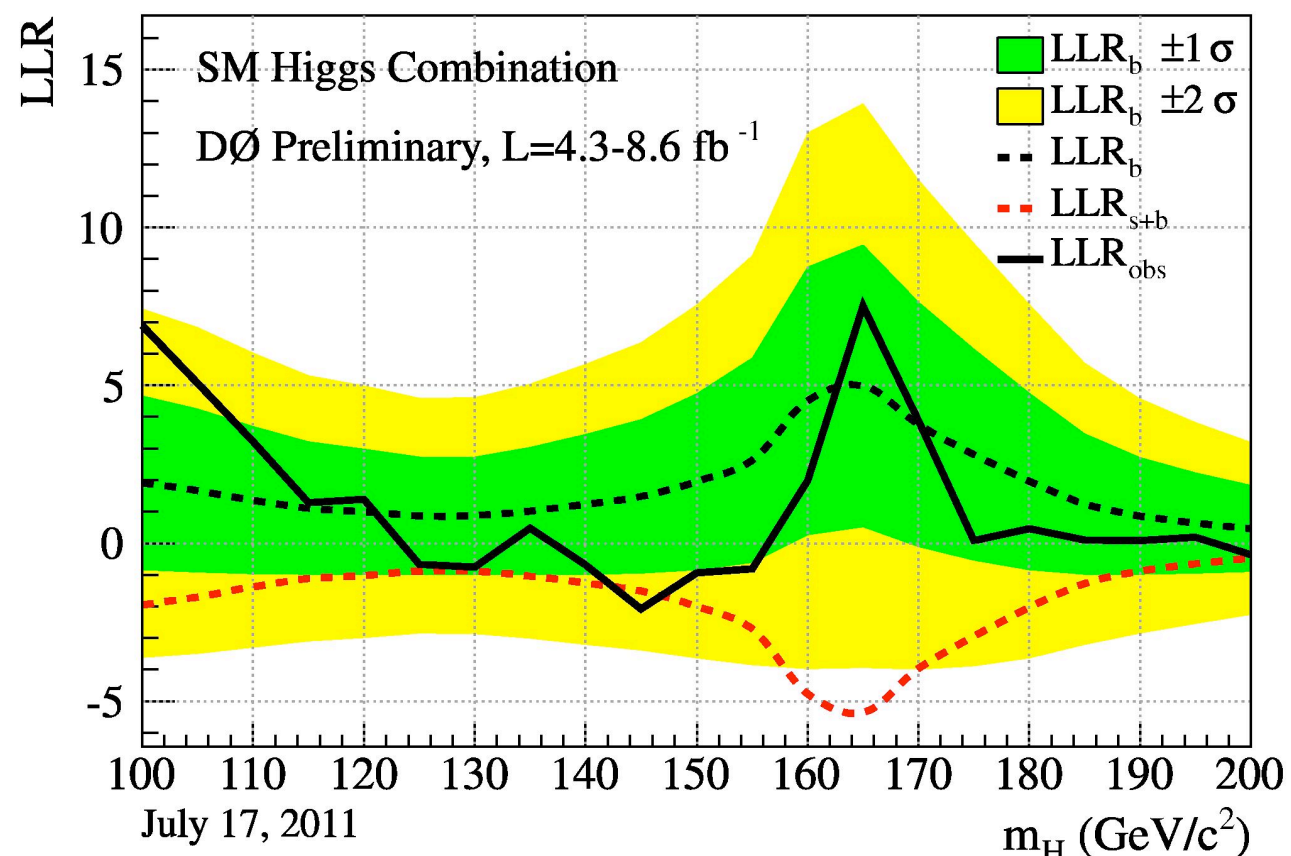


Relatively flat over the full mass range

For $m_H=110, 130, 160$ GeV observed limit = 20, 24, 11 $\times \sigma_{SM}$
compared to the expected limit of 14, 20, 12 $\times \sigma_{SM}$



Combining all 9 analyses:



Exclusion 161-170 GeV observed (159-170 GeV expected)

115 GeV : Obs (Exp) limit of 1.83 (1.9)

165 GeV : Obs (Exp) limit of 0.71 (0.87)

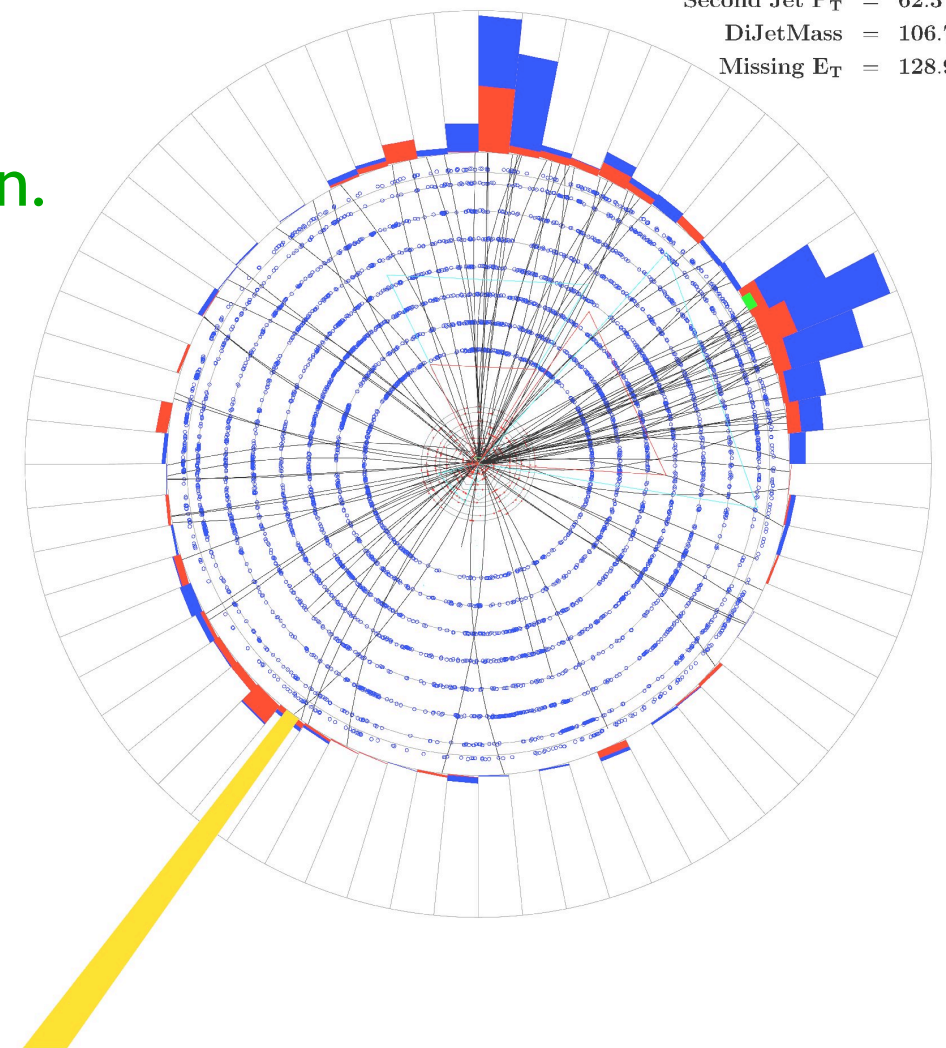
The DØ Higgs search program has excluded a high mass region and is rapidly approaching sensitivity at low mass.

With a well understood detector can use hadronically decaying taus to widen the search.

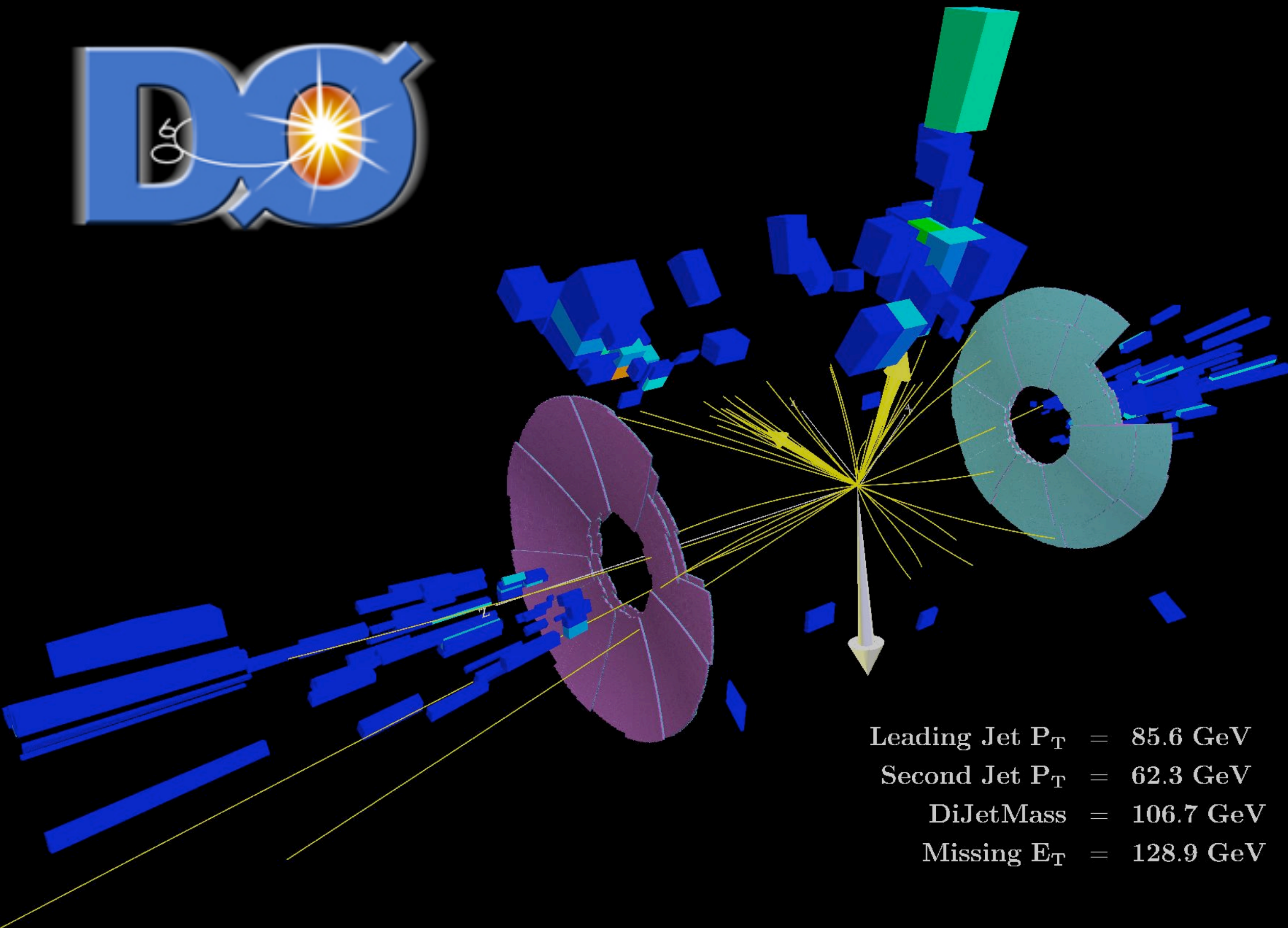
With the final data set and a long list of improvements this is going to be an exciting Winter conference season.

Run 248968 Evt 48062268 Fri Jan 23 06:59:26 2009

Leading Jet P_T = 85.6 GeV
Second Jet P_T = 62.3 GeV
DiJetMass = 106.7 GeV
Missing E_T = 128.9 GeV



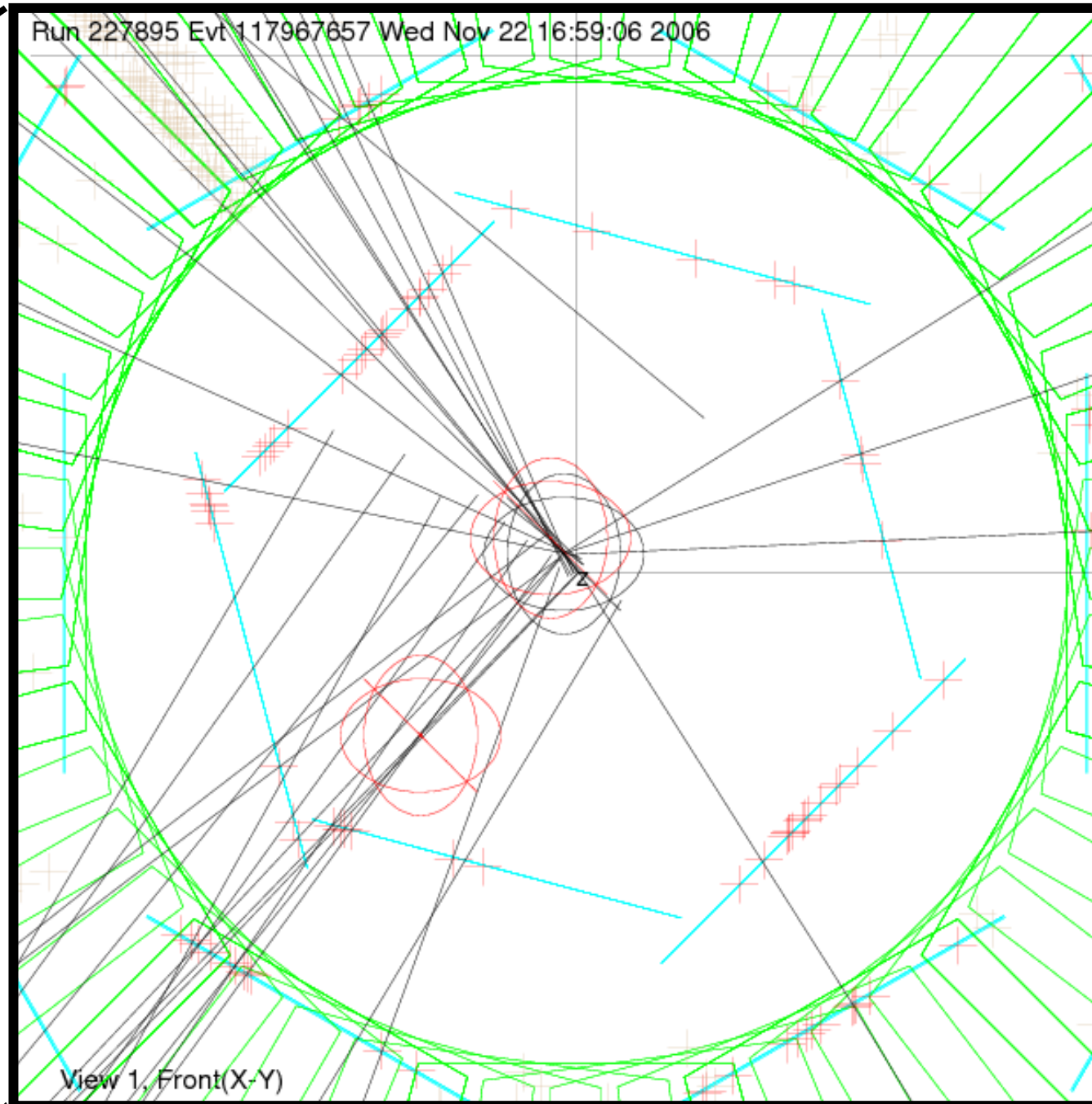
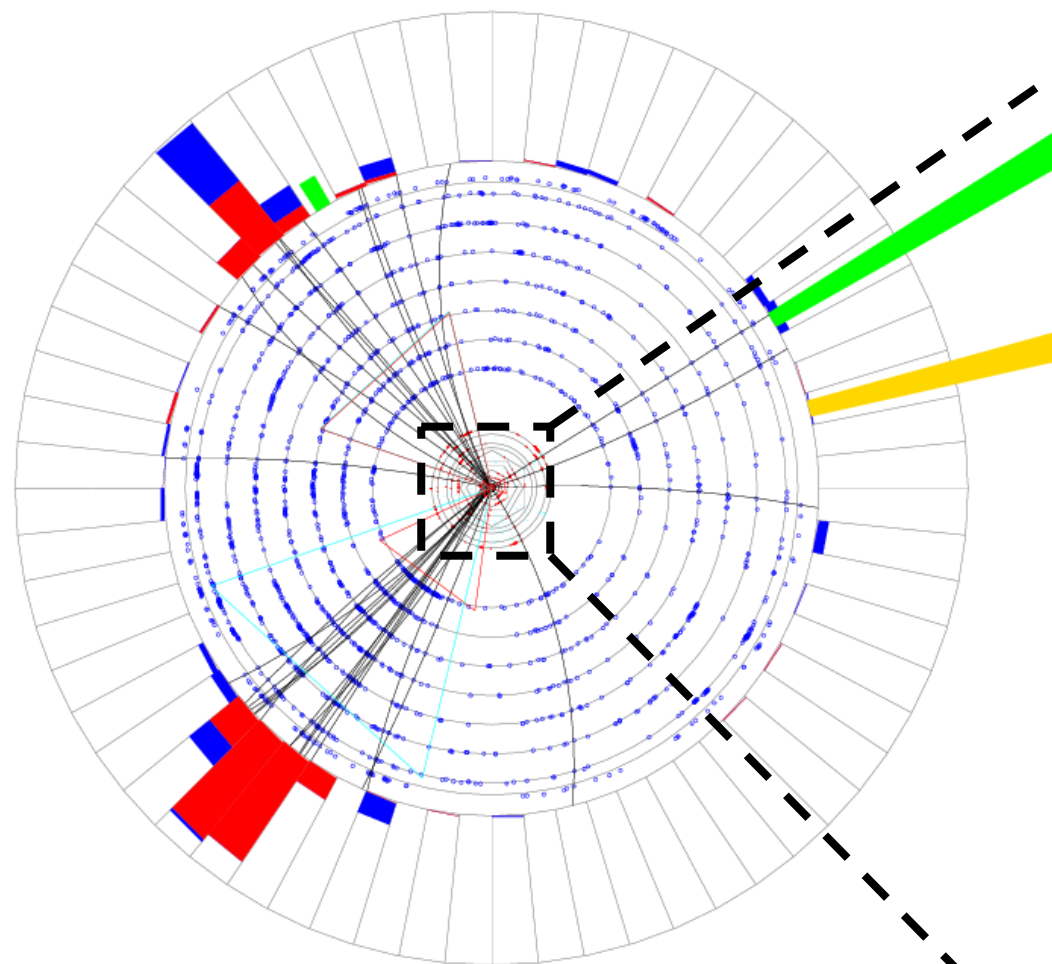
THANK YOU



Leading Jet P_T = 85.6 GeV
Second Jet P_T = 62.3 GeV
DiJetMass = 106.7 GeV
Missing E_T = 128.9 GeV

Run 227895 Evt 117967657 Wed Nov 22 16:59:06 2006

ET scale: 18 GeV



Vertex Tagging
(transverse plane)

(Signed) Track
Impact Parameter (dca)

Hard Scatter

Decay
Length (L_{xy})

displaced secondary vertex

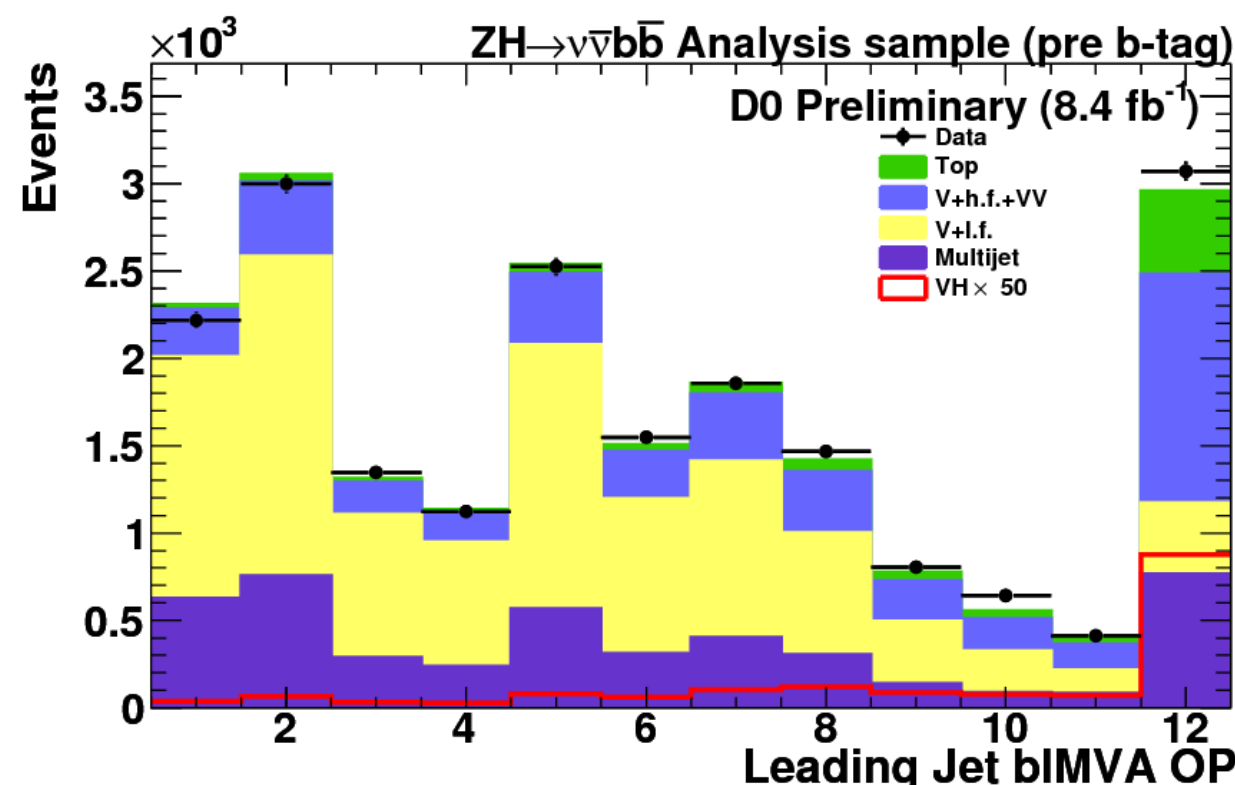
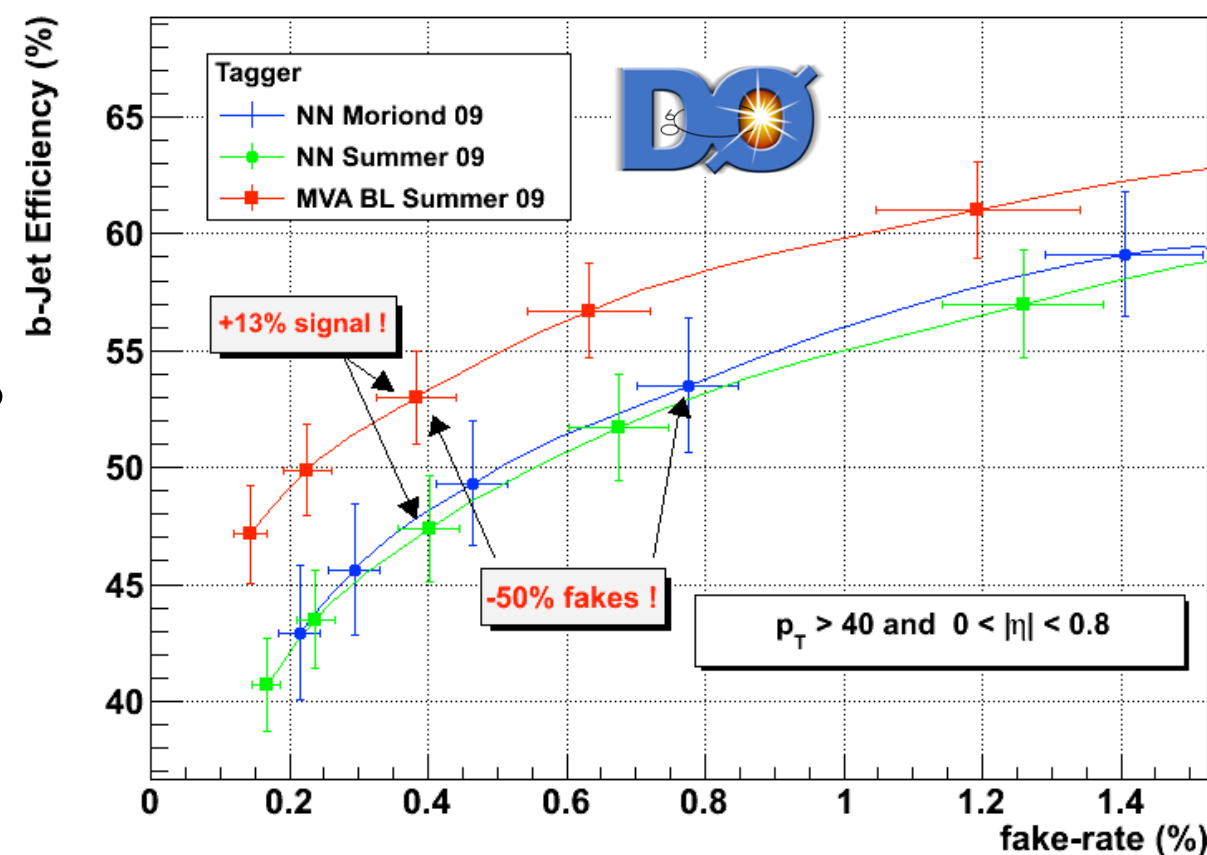
- MVA tagger
 - Better performance
- Modeling
 - Update on TRF, Fake rate measurement
 - Systematic uncertainty reduced by 50% on fake rate.
- Usage
 - Application of TRF
 - Use all operating point.

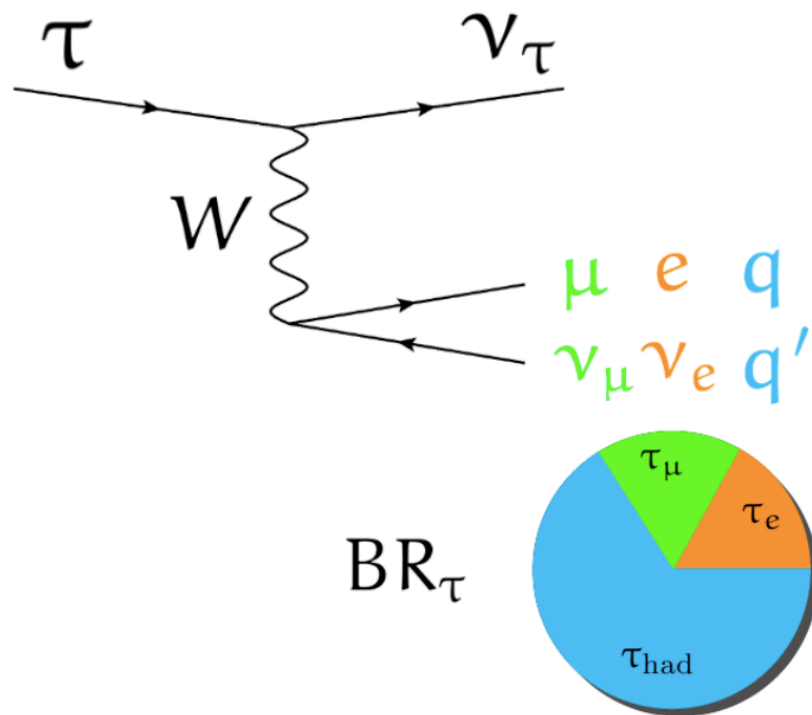
Use shape of bID MVA output in the final MVA

Two orthogonal sample

2 b-tag: both jet pass Loosest tag

1 b-tag: one of jet pass Loosest tag





Hadronic Taus can be put into three groups based on typical decay mode signatures:

Type 1: 1 track + Energy in hadronic calorimeter

Corresponds mostly to:

$$\tau^\pm \rightarrow \pi^\pm \nu_\tau$$

Type 2: 1 track + Energy in hadronic calorimeter + 1 associated EM subcluster

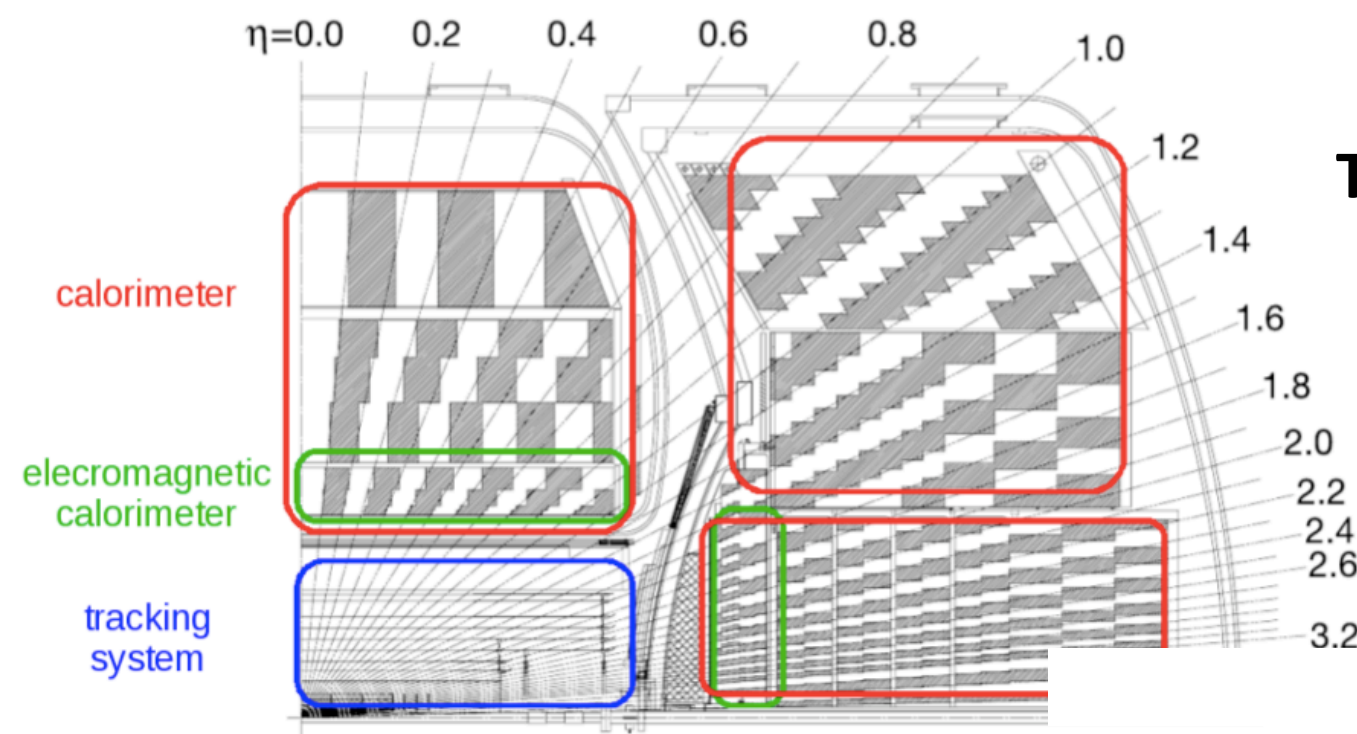
Corresponds mostly to:

$$\tau^\pm \rightarrow \rho^\pm (\rightarrow \pi^0 \pi^\pm) \nu_\tau$$

Type 3: 2 or more tracks + Energy in hadronic calorimeter

Corresponds mostly to:

$$\tau^\pm \rightarrow a_1^\pm (\rightarrow \rho^0 \pi^\pm \rightarrow 3\pi^\pm) \nu_\tau$$



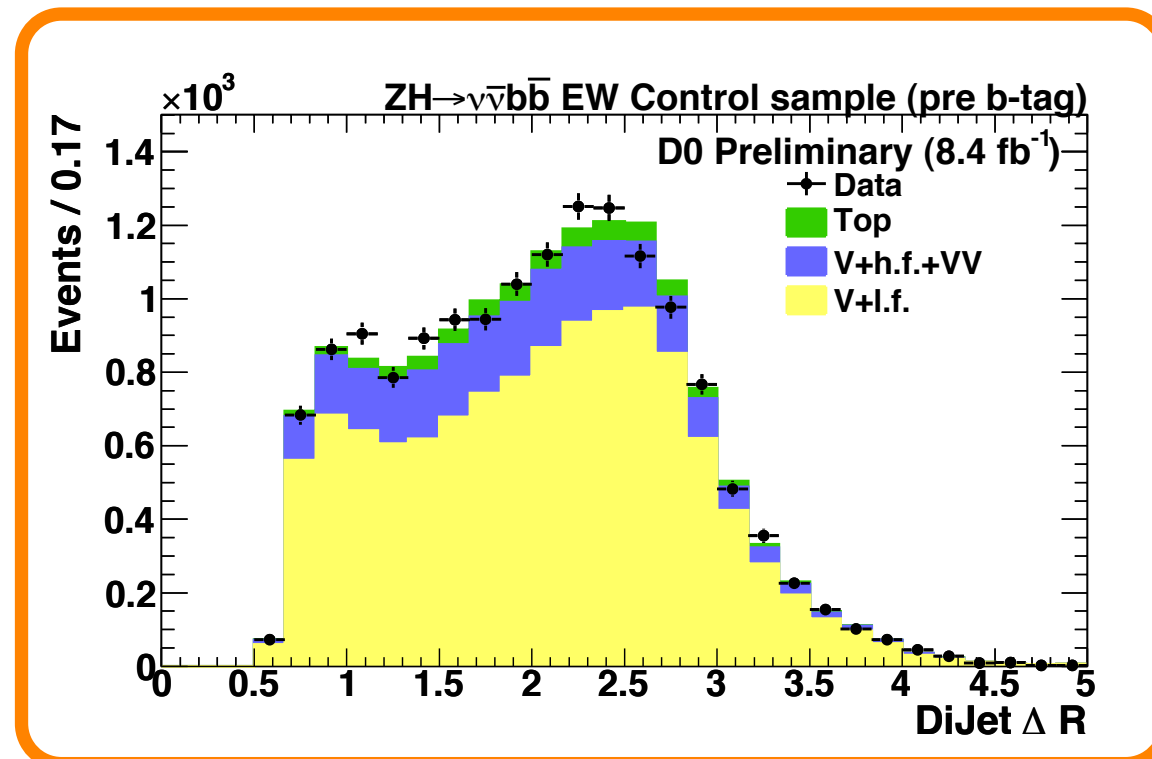
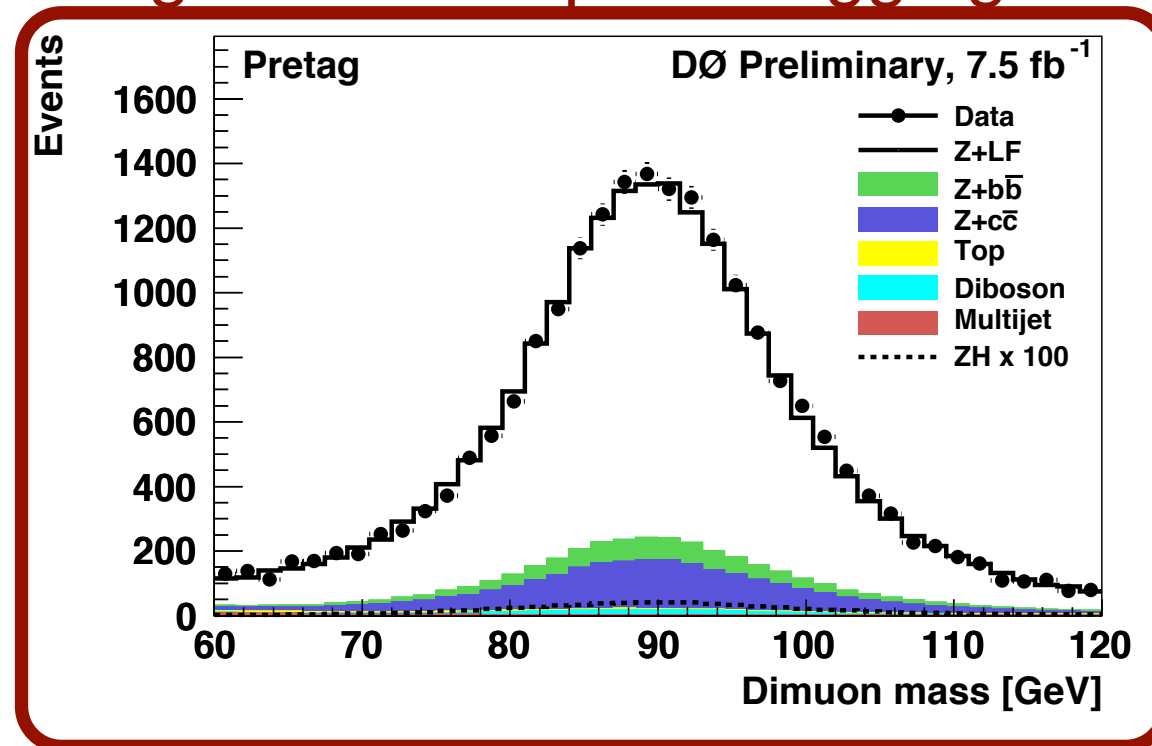
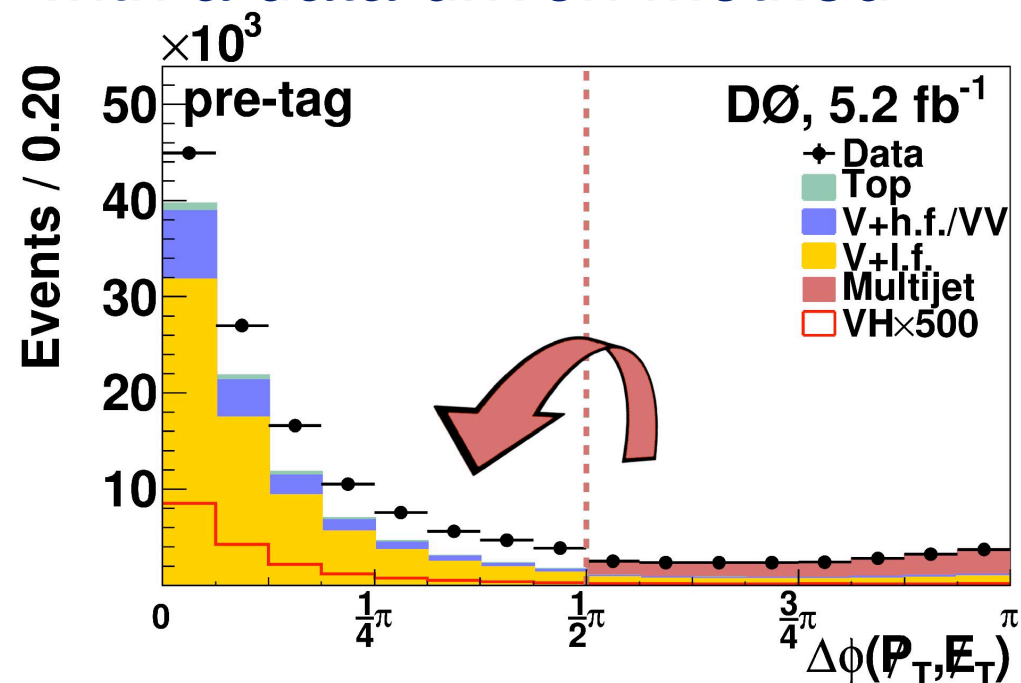
Major backgrounds are V +jets, $t\bar{t}$, and multijet production (MJ)

Control samples used to validate modeling with **high-statistics pre-b-tagging** samples or **alternate selection**:

$ZH \rightarrow \nu\nu b\bar{b}$: requires a muon(EW)
loosens cuts (MJ)

$H+X$: invert τ NN & $MET > 30$ GeV (W+jets)
replacing the τ req with muon (Z+jet)

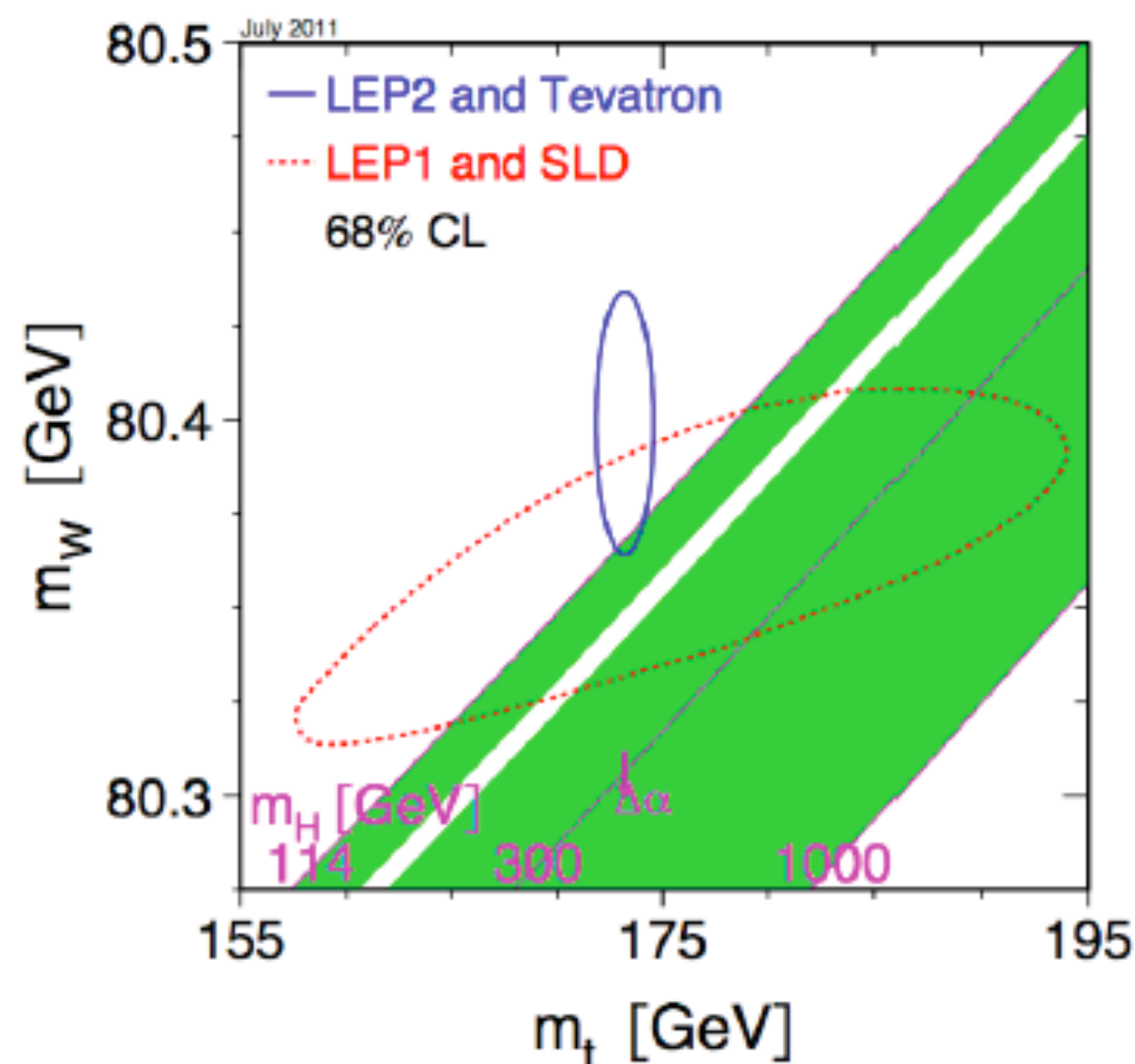
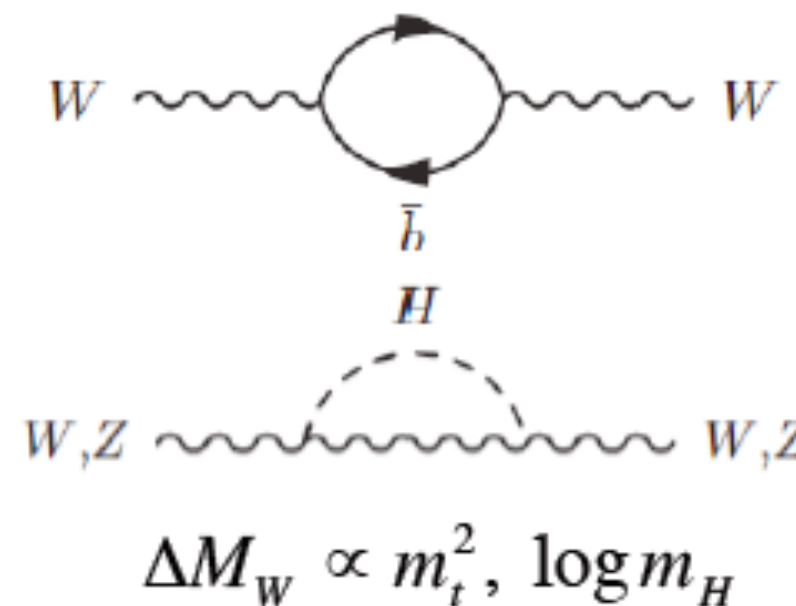
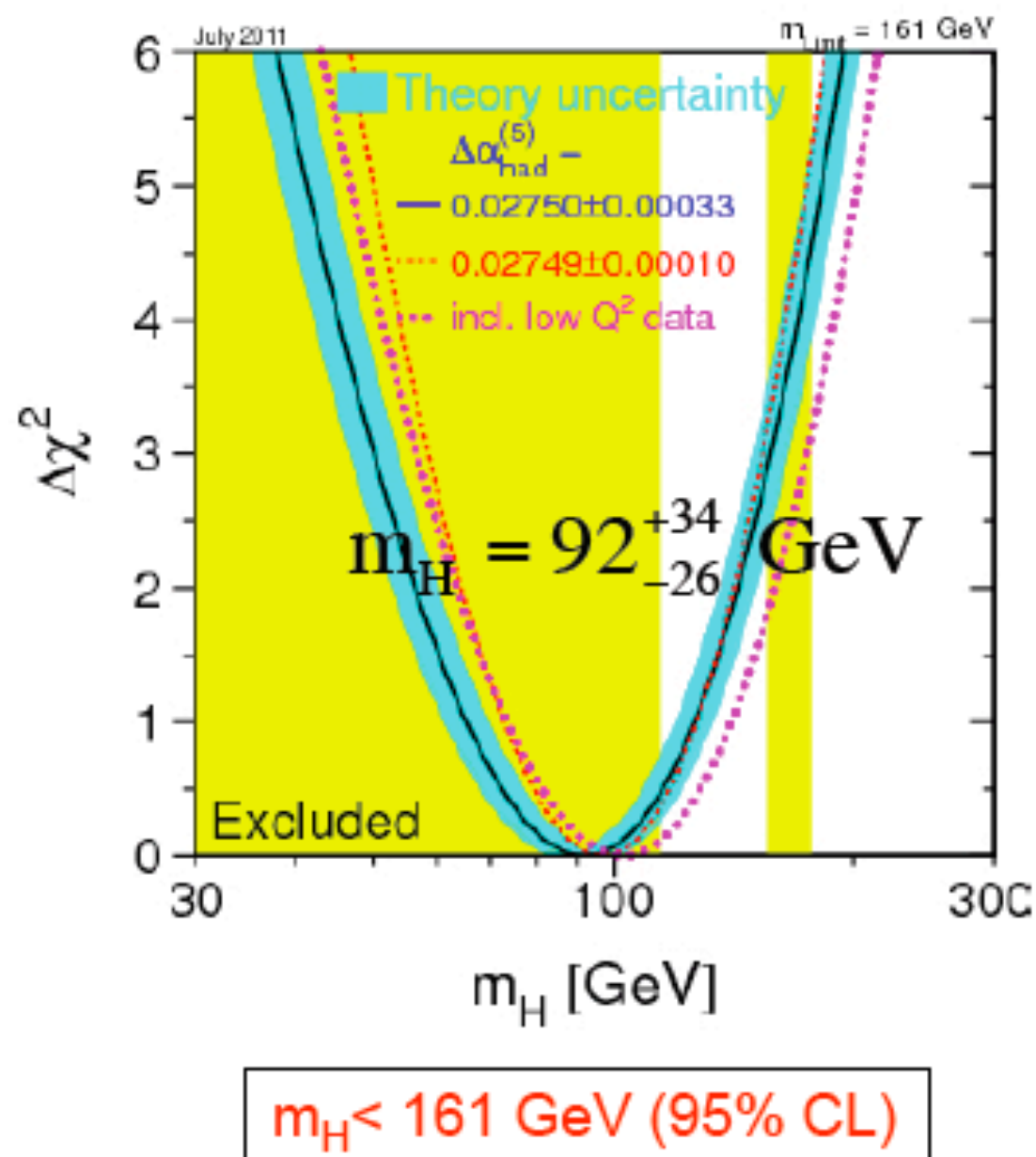
Multijet background determined with a data driven method

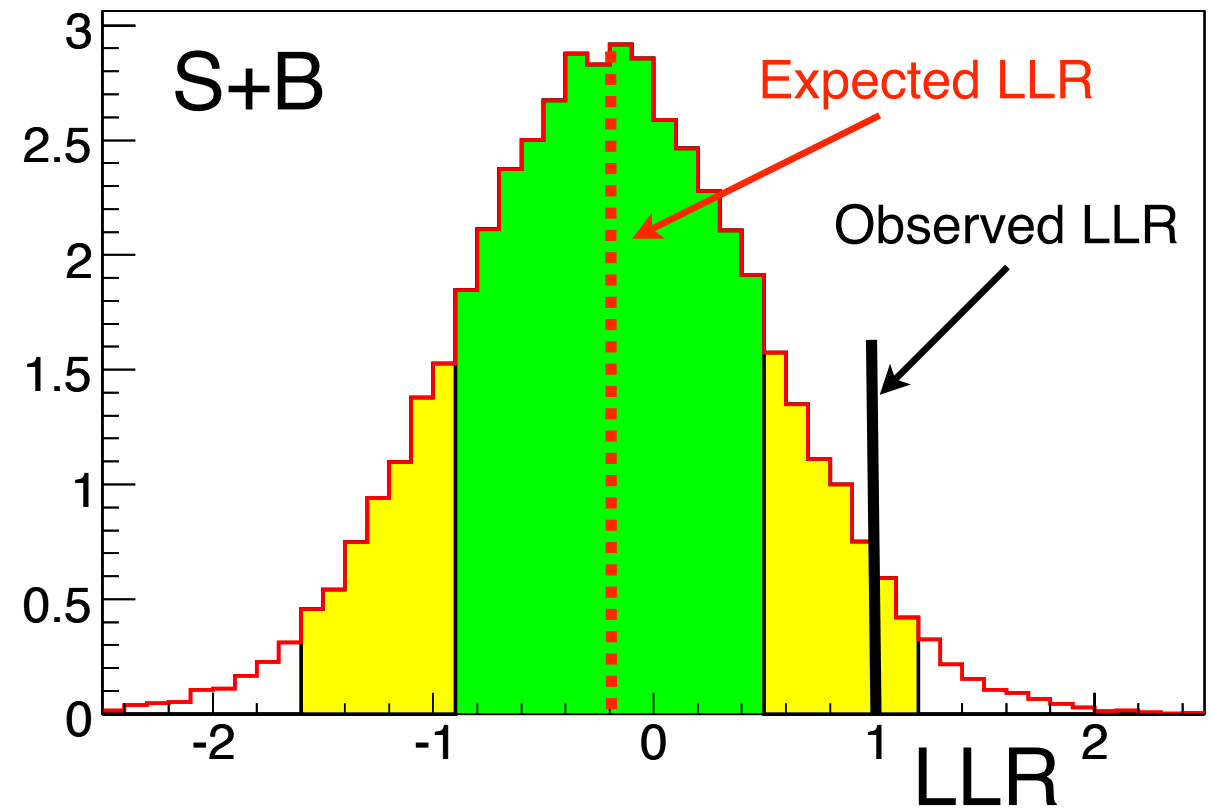
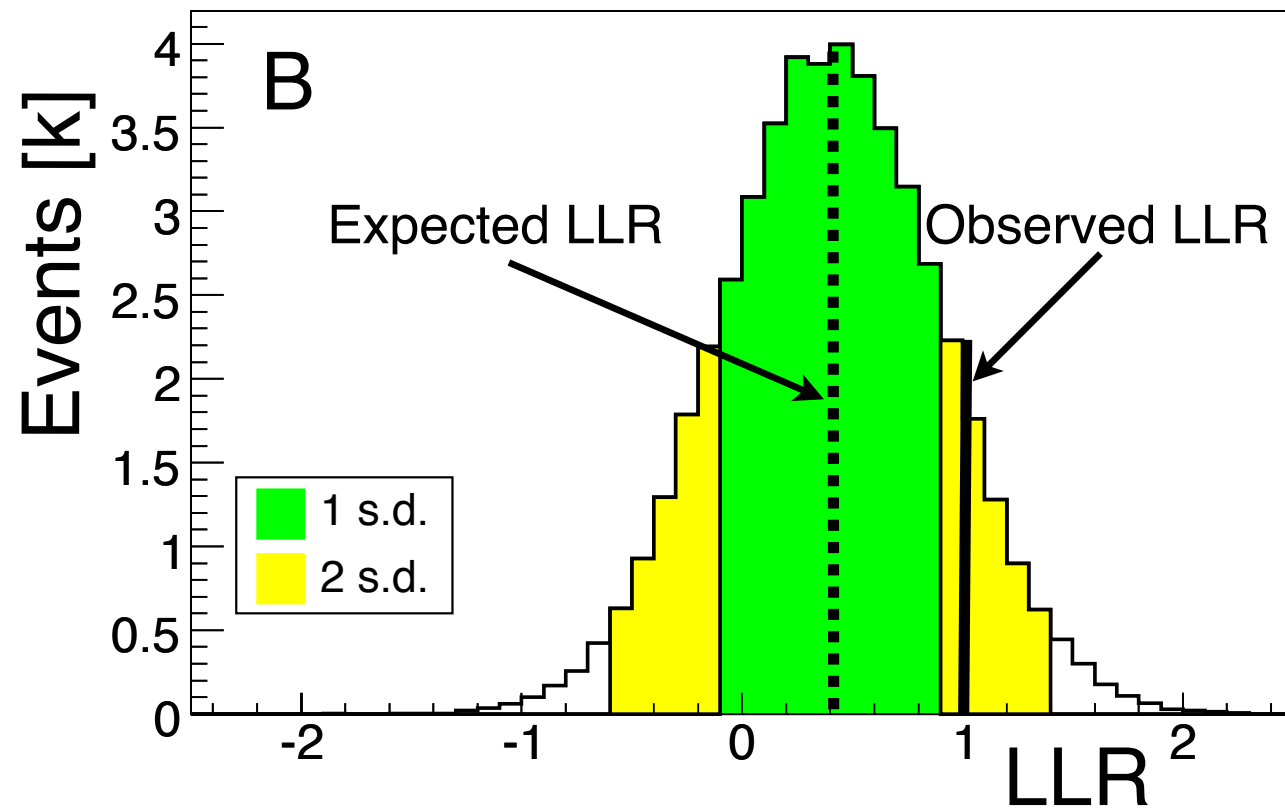


Indirect Constraints

Indirect constraints

- Precision electroweak observables are sensitive to the Higgs boson mass via quantum corrections.

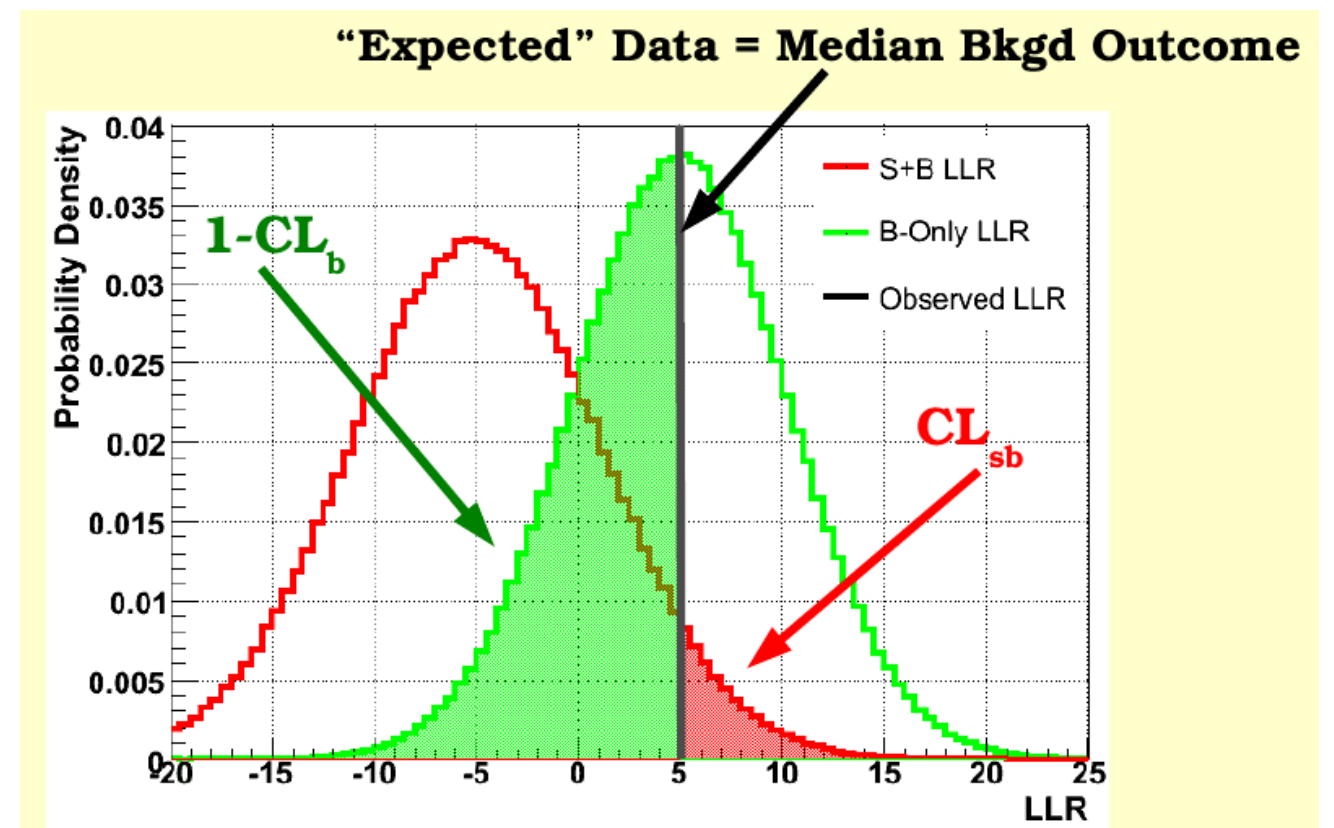
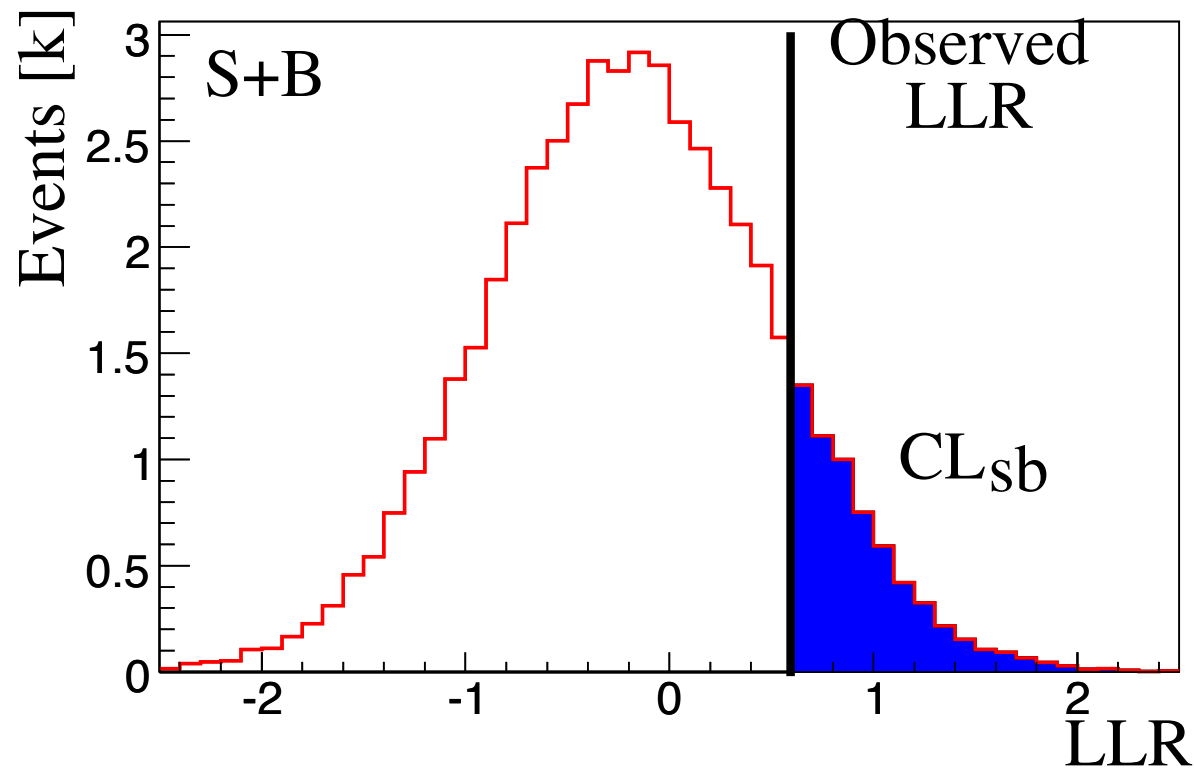




- Repeat calculation but with pseudo-data obtained by a Poisson fluctuation of b_i in each bin (B) or $s_i + b_i$ in each bin (S+B).
- Repeat many times to obtain LLR distribution: median is **Expected LLR**
- Nuisance parameters (systematics) alter prediction: $b_i \rightarrow b_i + \theta_i$.
- Smear pseudo-experiments by Gaussian priors for nuisance parameters.
- Fit systematics: $L(S+B)/L(B) \rightarrow L(S+B, \theta_{S+B}^{\text{FIT}})/L(B, \theta_B^{\text{FIT}})$

$$\chi^2 = -2\ln L = 2 \sum_{i=0}^{N_{\text{bins}}} (\hat{B}_i - D_i) - D_i \ln \left(\frac{\hat{B}_i}{D_i} \right) + \sum_{k=0}^{N_{\text{syst}}} S_k^2 \quad \hat{B}_i \rightarrow B_i \prod_{k=0}^{N_{\text{syst}}} (1 + \sigma_i^k S_k)$$

B_i = nominally predicted bin content
 σ_i^k = fractional uncertainty
 S_k = N sigma deviation from nominal



Significance (for cross sections p-values & CLs have same interpretation)

- CL is fraction of pseudo-experiments with LLR above Observed LLR
- How likely is a pseudo-experiment to be more background-like then observed?
- $CL_s = CL_{sb} / CL_b$ is safe against highly unlikely B-like fluctuate
- CL_s over-covers CL_b (95% CL with $CL_s \sim 98\%$ CL with CL_b)

Cross Section

- Limit - Find value with satisfies the exclusion condition
- Measurement - Float signal in κ^2 minimization

The input channels.

TABLE I: List of analysis channels, corresponding integrated luminosities, and final variables used for setting limits, which is either a decision-tree (DTree) or neural-network (NN) discriminant. See Sect. I for details ($\ell = e, \mu$).

Channel	Luminosity (fb^{-1})	Final Variable	# Sub-Channels	Data Epochs	Reference
$WH \rightarrow \ell \nu b \bar{b}$, ST/DT, 2/3 jet	8.5	DTree discriminant	8	3	[4]
$ZH \rightarrow \nu \bar{\nu} b \bar{b}$, ST/DT	8.4	DTree discriminant	2	3	[5]
$ZH \rightarrow \ell \ell b \bar{b}$, ST/DT	8.6	DTree discriminant	10	3	[6]
$H \rightarrow W^+ W^- \rightarrow \ell^\pm \nu \ell^\mp \nu$, 0/1/2+ jet	8.1	DTree discriminant	9	2	[7]
$H \rightarrow W^+ W^- \rightarrow \ell \nu q \bar{q}$	5.4	DTree discriminant	2	2	[8]
$H + X \rightarrow \mu^\pm \tau_{had}^\mp + \leq 1j$	7.3	NN discriminant	3	1	[9]
$VH \rightarrow \ell^\pm \ell^\pm + X$	5.3	DTree discriminant	3	2	[10]
$H + X \rightarrow \ell^\pm \tau_{had}^\mp jj$	4.3	DTree discriminant	2	1	[11]
$H \rightarrow \gamma \gamma$	8.2	DTree discriminant	1	1	[12]

[link to diphoton seach TeV combo - Xuebing Bu](#)

II. SIGNAL PREDICTIONS AND UNCERTAINTIES

A common approach to the signal predictions and associated uncertainties is followed by the CDF and DØ Collaborations. An outline of the procedures followed is given here; more complete discussion can be found in Ref. [20].

The Monte Carlo signal simulation is provided by the LO generator PYTHIA (with CTEQ5L and CTEQ6L1 [38] leading-order (LO) parton distribution functions) which includes a parton shower and fragmentation and hadronization models. We reweight the Higgs boson p_T spectra in our PYTHIA Monte Carlo samples to that predicted by HQT [39] when making predictions of differential distributions of GGF signal events. To evaluate the impact of the scale uncertainty on our differential spectra, we use the RESBOS [40] generator, and apply the scale-dependent differences in the Higgs boson p_T spectrum to the HQT prediction, and propagate these to our final discriminants as a systematic uncertainty on the shape, which is included in the calculation of the limits.

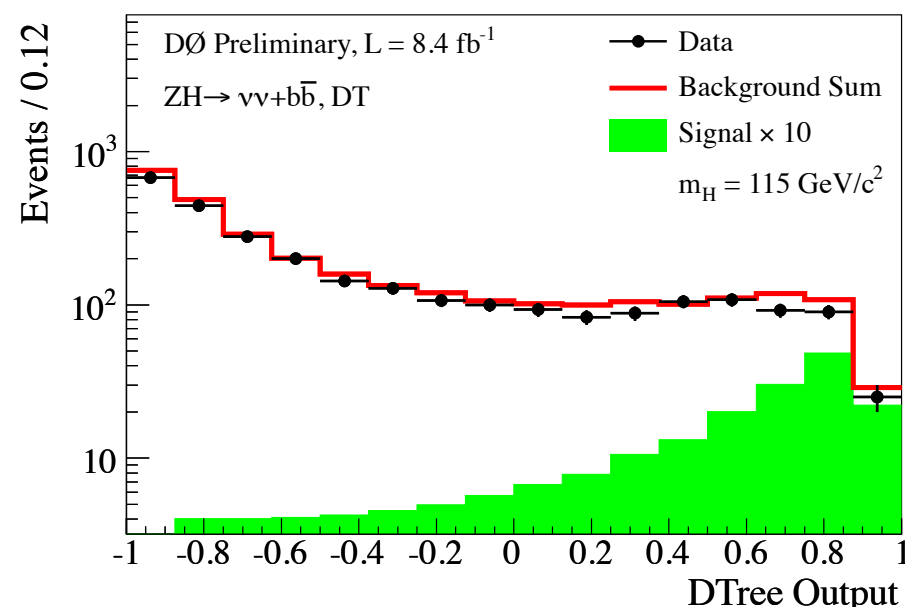
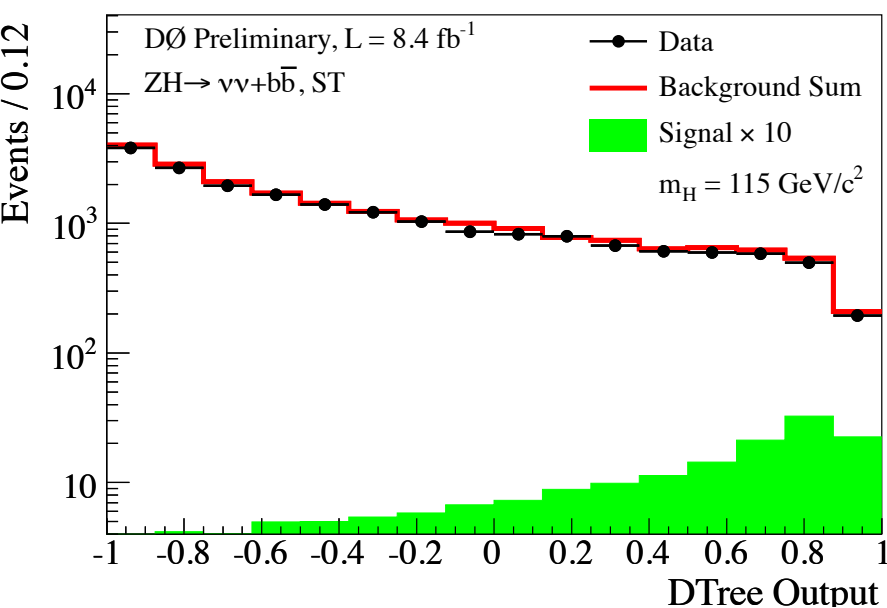
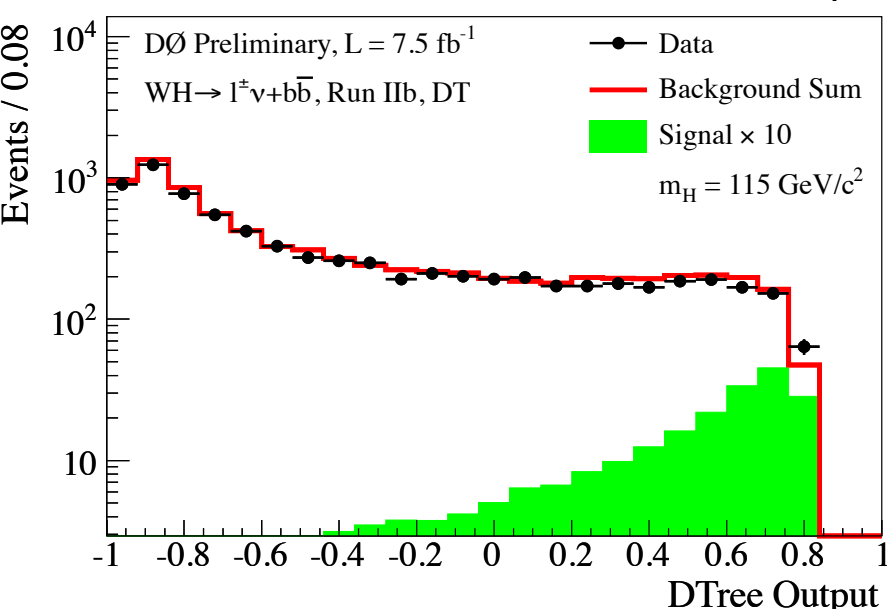
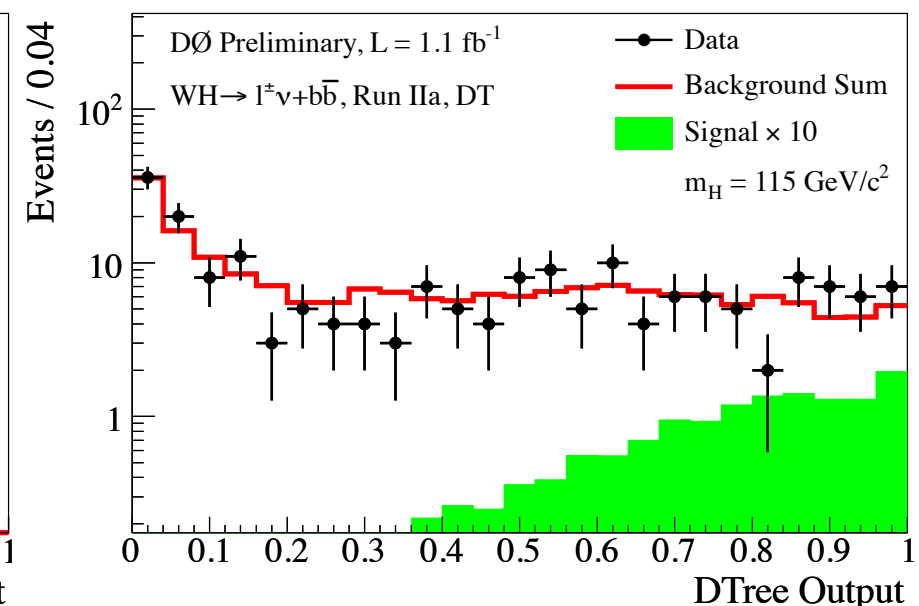
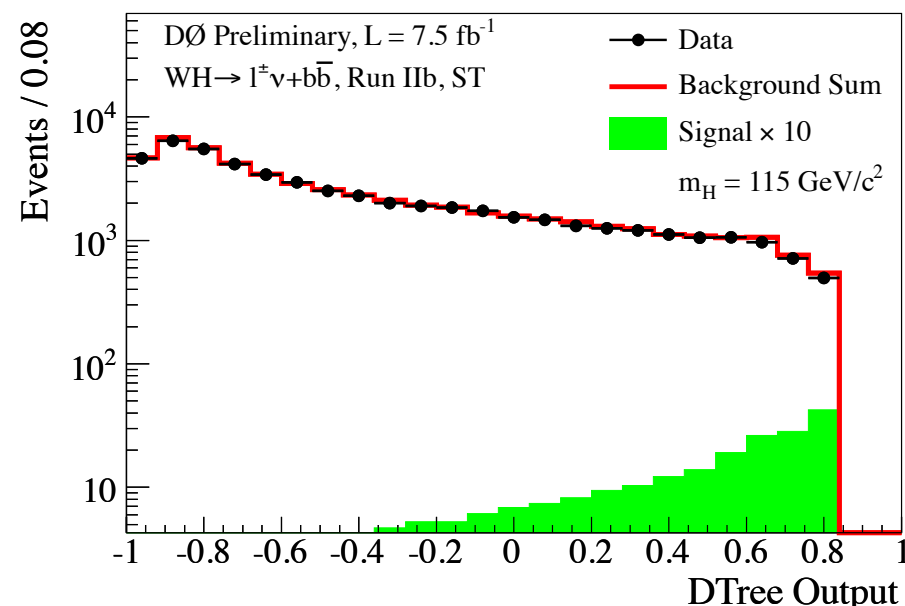
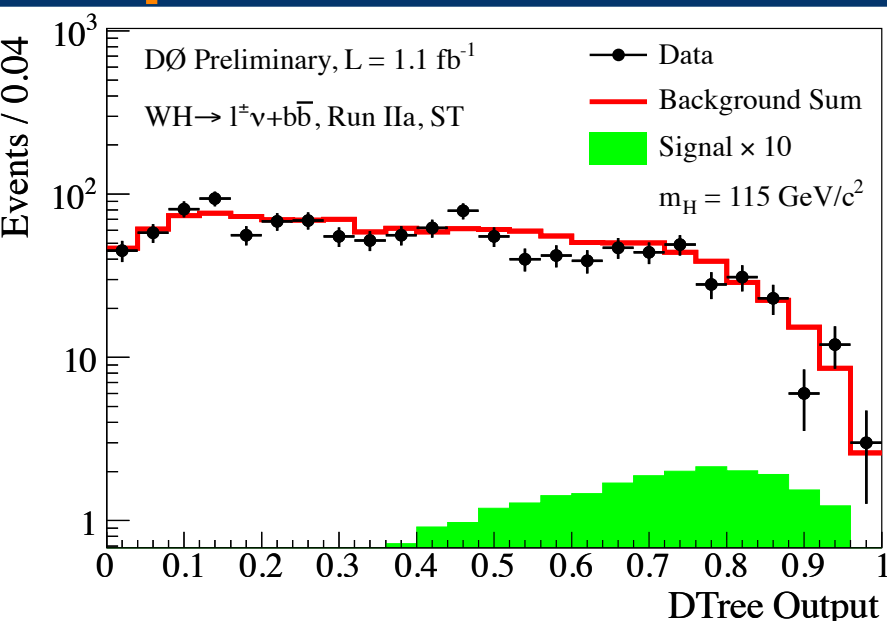
We normalize our Higgs boson signal predictions to the most recent high-order calculations available. The $gg \rightarrow H$ production cross section we use is calculated at next-to-next-to leading order (NNLO) in QCD with a next-to-next-to leading log (NNLL) resummation of soft gluons; the calculation also includes two-loop electroweak effects and handling of the running b quark mass [21, 22]. The numerical values in Table II are updates [23] of these predictions with m_t set to 173.1 GeV/ c^2 [24], and an exact treatment of the massive top and bottom loop corrections up to NLO + next-to-leading-log accuracy. The factorization and renormalization scale choice for this calculation is $\mu_F = \mu_R = m_H$. These calculations are refinements of the earlier NNLO calculations of the $gg \rightarrow H$ production cross section [26–28]. Electroweak corrections were computed in Refs. [29, 30]. Soft gluon resummation was introduced in the prediction of the $gg \rightarrow H$ production cross section in Ref. [31]. The $gg \rightarrow H$ production cross section depends strongly on the gluon parton density function, and the accompanying value of $\alpha_s(q^2)$. The cross sections used here are calculated with the MSTW 2008 NNLO PDF set [32], as recommended by the PDF4LHC working group [33]. The inclusive Higgs boson production cross sections are listed in Table II.

For analyses that consider inclusive $gg \rightarrow H$ production but do not split it into separate channels based on the number of reconstructed jets, we use the inclusive uncertainties from the simultaneous variation of the factorization and renormalization scale up and down by a factor of two. We use the prescription of the PDF4LHC working group for evaluating PDF uncertainties on the inclusive production cross section. QCD scale uncertainties that affect the cross section via their impacts on the PDFs are included as a correlated part of the total scale uncertainty. The remainder of the PDF uncertainty is treated as uncorrelated with the QCD scale uncertainty.

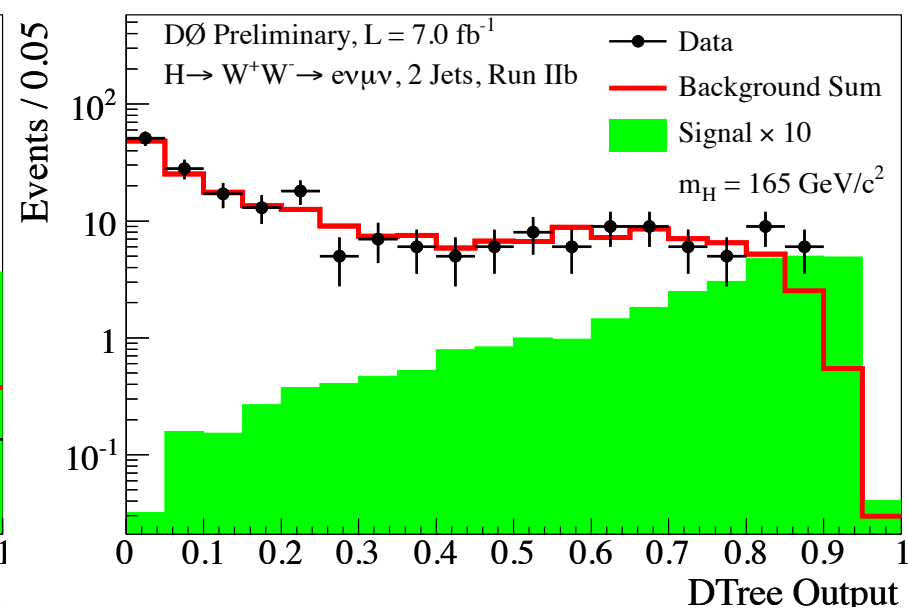
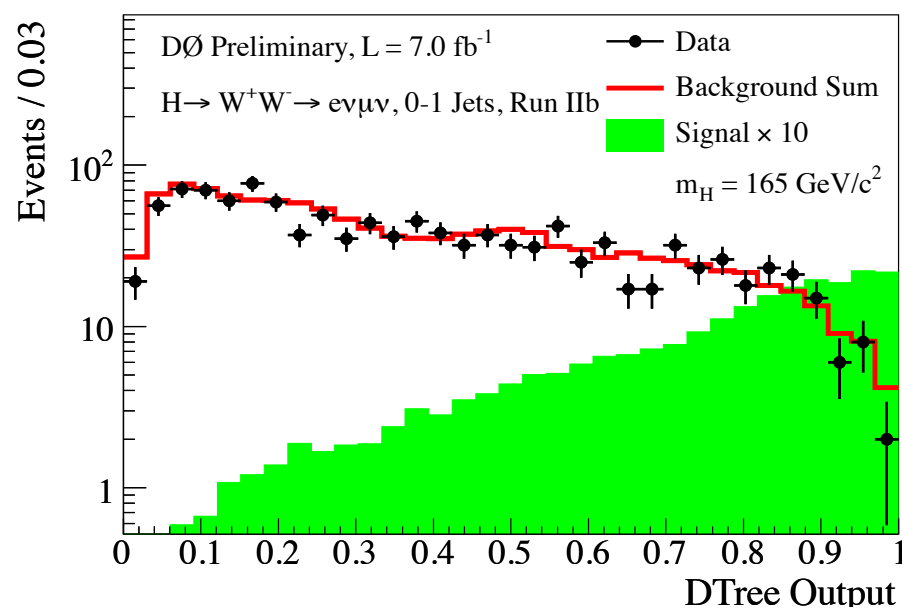
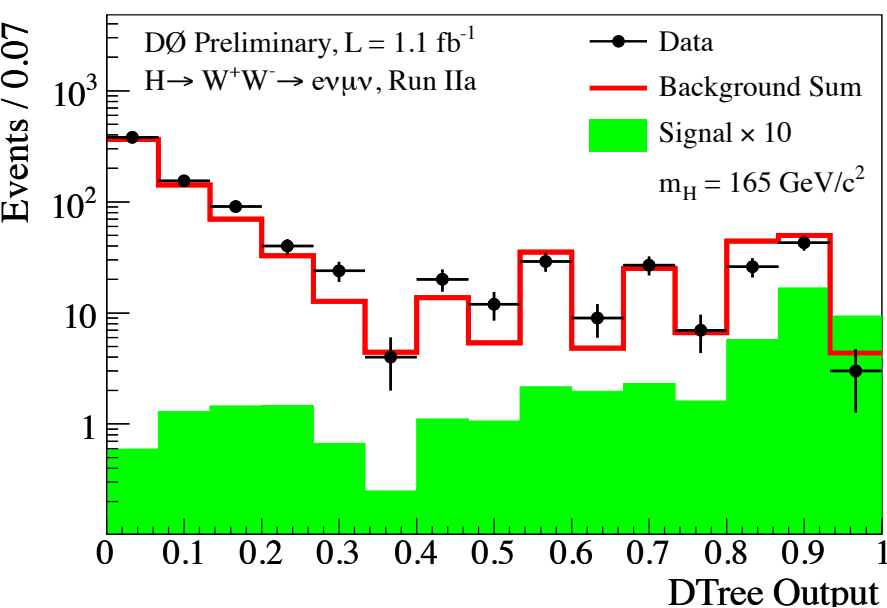
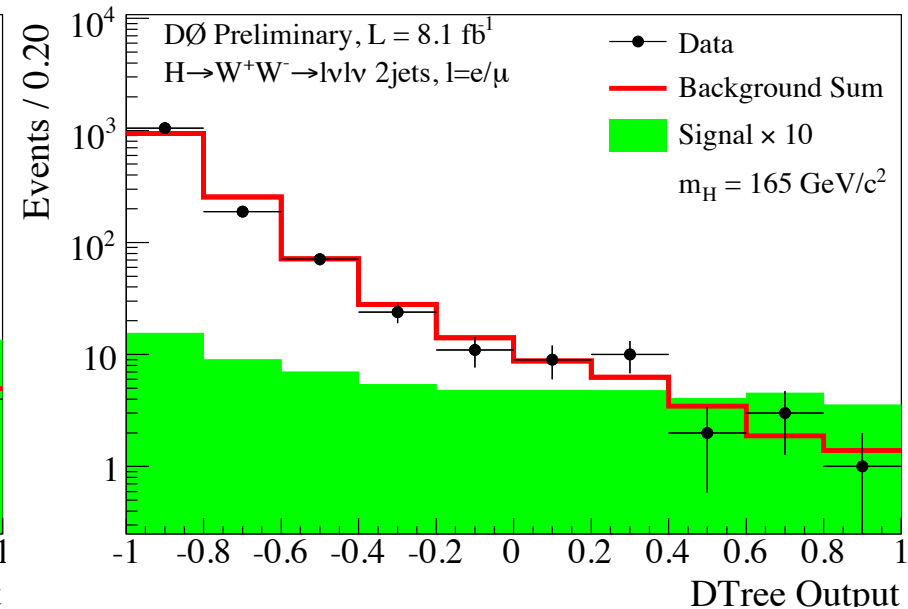
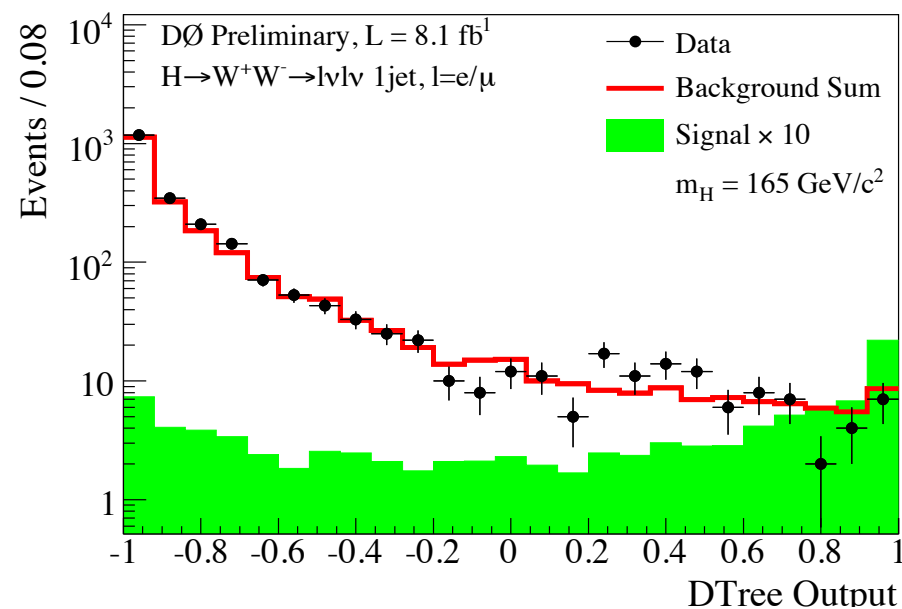
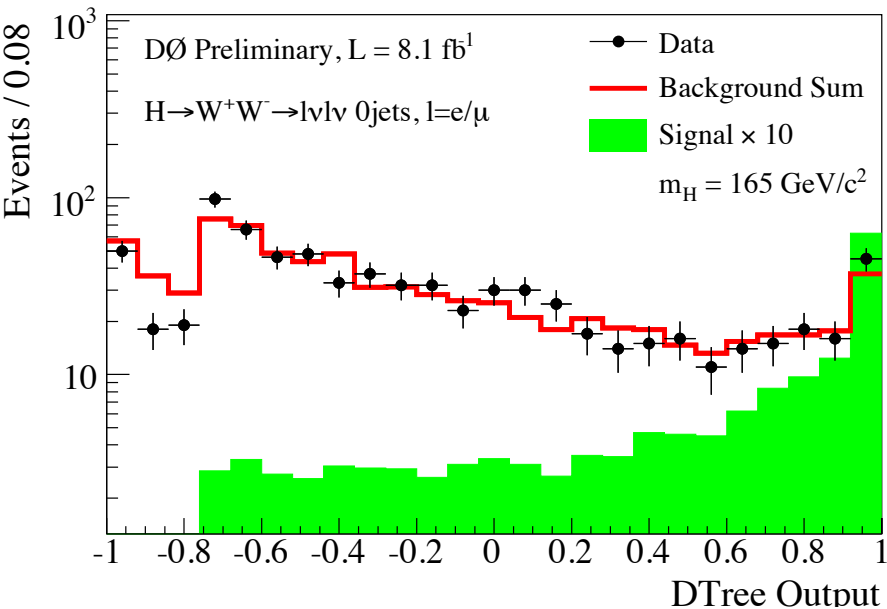
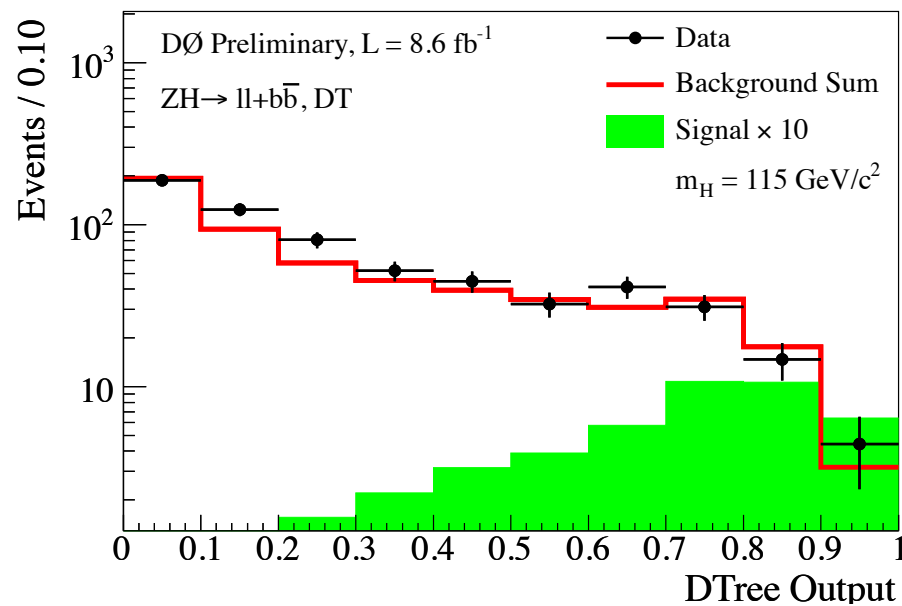
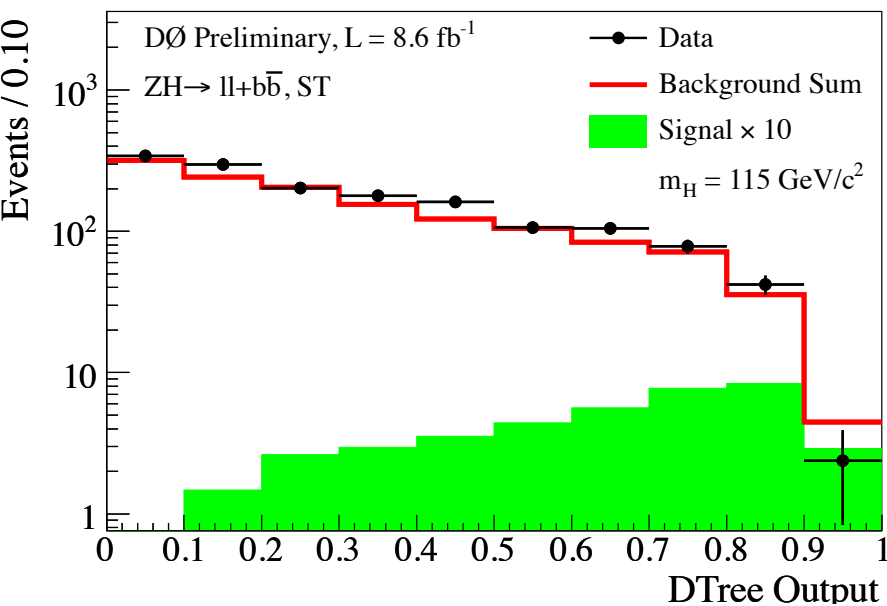
For analyses seeking $gg \rightarrow H$ production that divide events into categories based on the number of reconstructed jets, we employ a new approach for evaluating the impacts of the scale uncertainties. Following the recommendations of Ref. [25], we treat the QCD scale uncertainties obtained from the NNLL inclusive [21, 22], NLO one or more jets [36], and NLO two or more jets [37] cross section calculations as uncorrelated with one another. We then obtain QCD scale uncertainties for the exclusive $gg \rightarrow H + 0$ jet, 1 jet, and 2 or more jet categories by propagating the uncertainties on the inclusive cross section predictions through the subtractions needed to predict the exclusive rates. For example, the $H+0$ jet cross section is obtained by subtracting the NLO $H + 1$ or more jet cross section from the inclusive NNLL+NNLO cross section. We now assign three separate, uncorrelated scale uncertainties which lead to correlated and anticorrelated uncertainty contributions between exclusive jet categories. The procedure in Ref. [36] is used to determine PDF model uncertainties. These are obtained separately for each jet bin and treated as 100% correlated between jet bins.

Another source of uncertainty in the prediction of $\sigma(gg \rightarrow H)$ is the extrapolation of the QCD corrections computed for the heavy top quark loops to the light-quark loops included as part of the electroweak corrections. Uncertainties at the level of 1-2% are already included in the cross section values we use [21, 22]. In Ref. [21], it is argued that the factorization of QCD corrections is known to work well for Higgs boson masses many times in excess of the masses of the loop particles. A 4% change in the predicted cross section is seen when all QCD corrections are removed from the diagrams containing light-flavored quark loops, which is too conservative. For the b quark loop, which is computed separately in Ref. [21], the QCD corrections are much smaller than for the top loop, further giving confidence that it does not introduce large uncertainties.

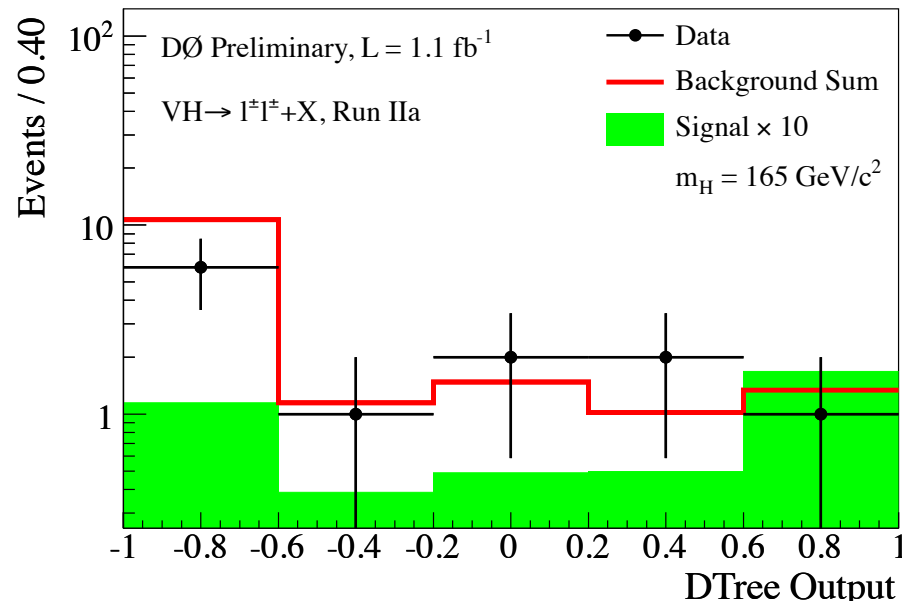
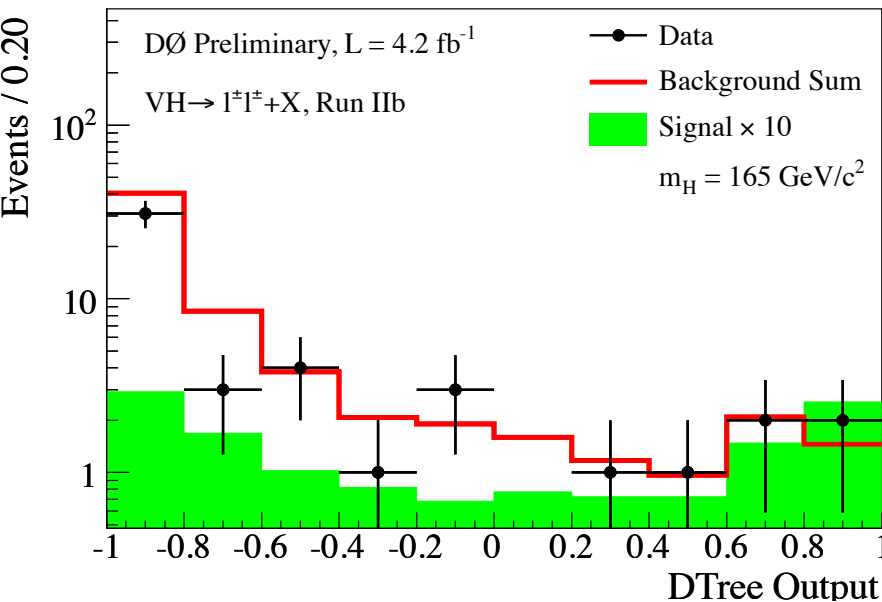
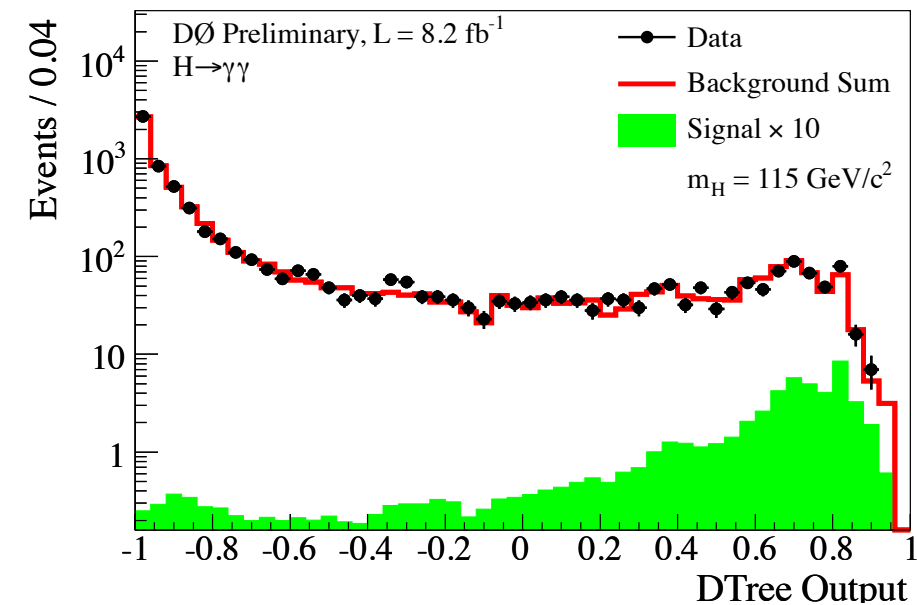
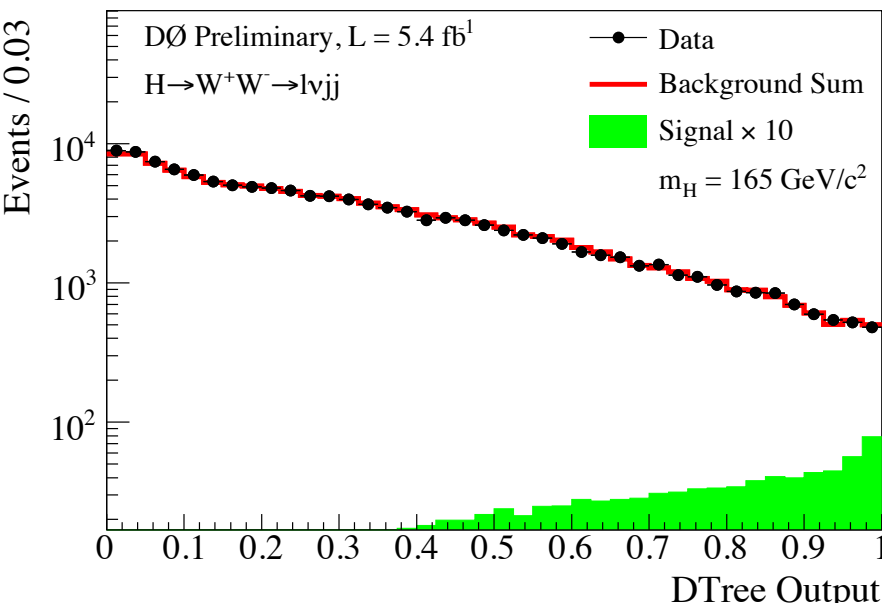
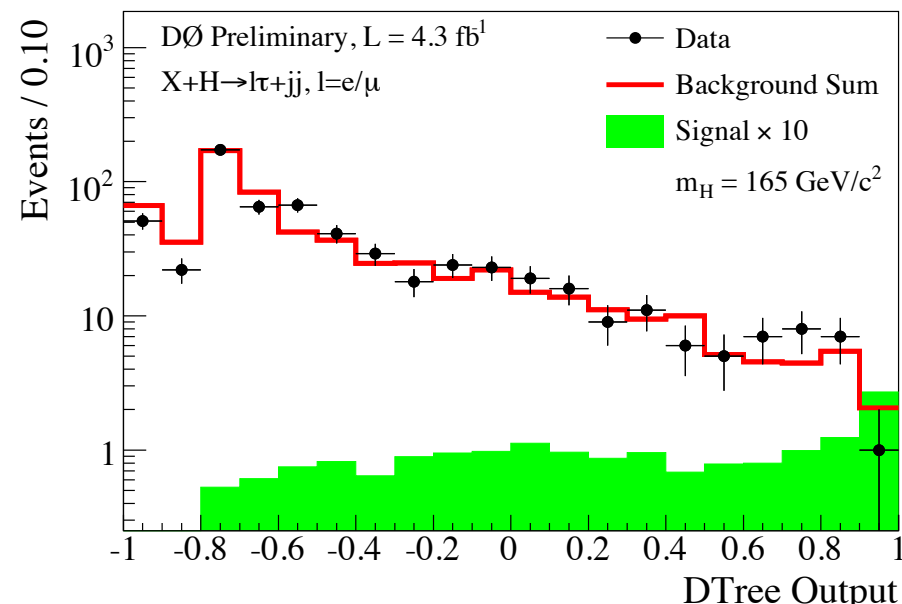
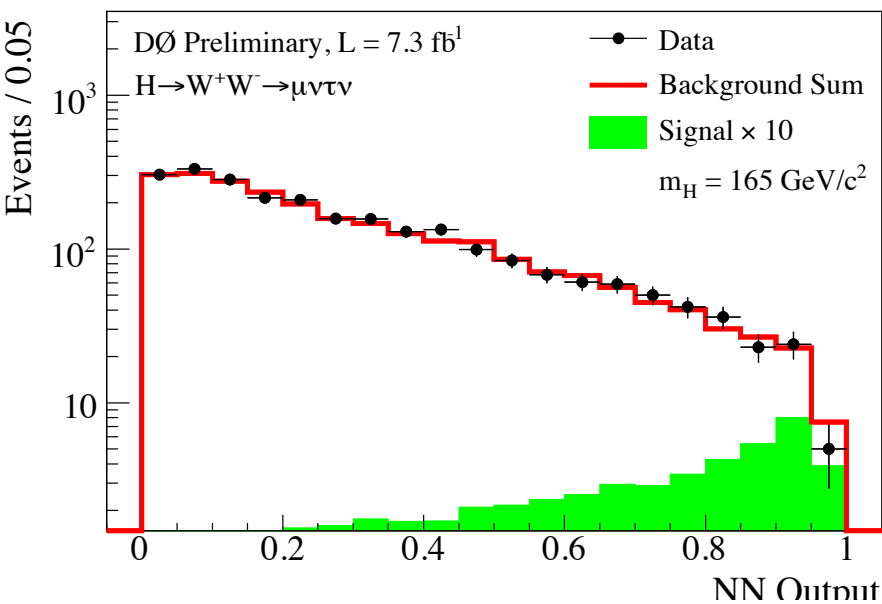
Inputs



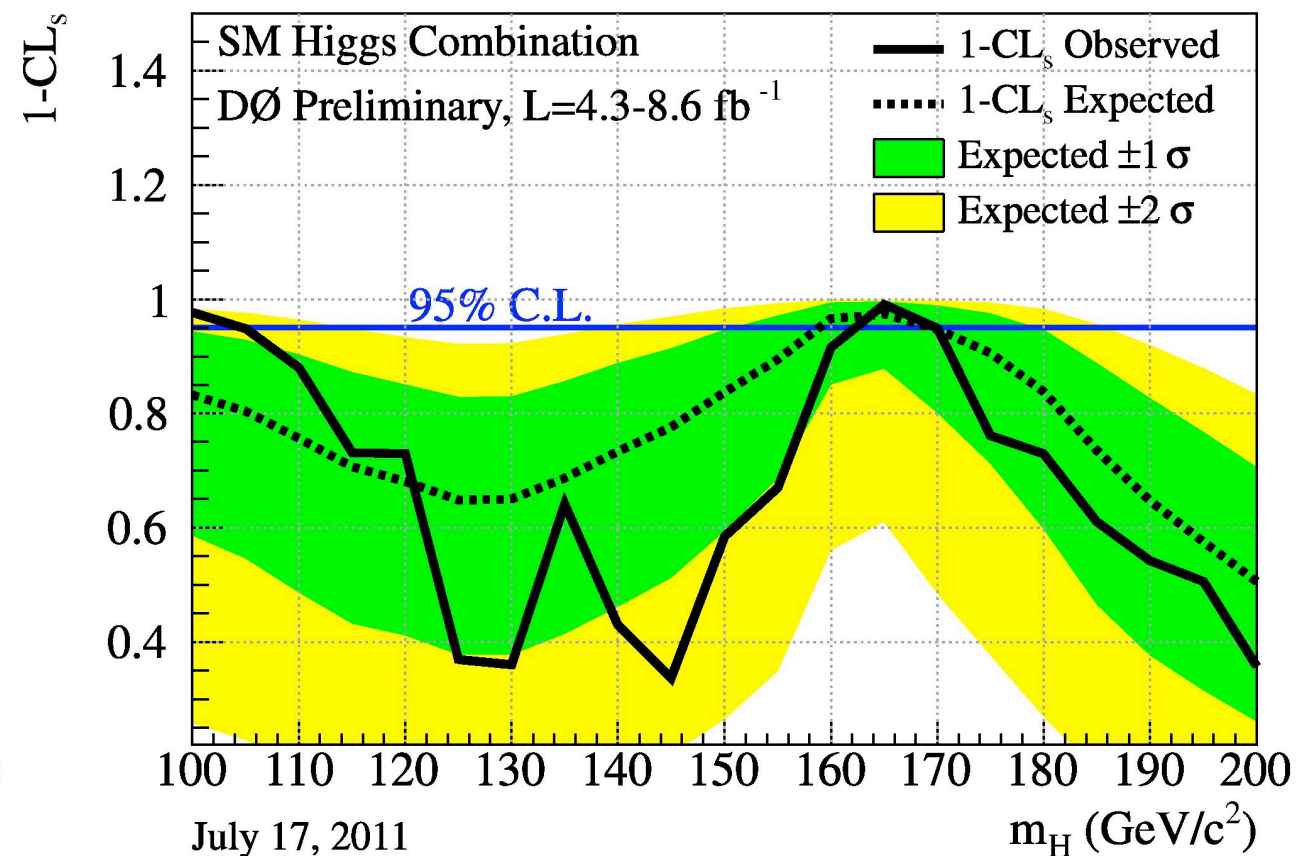
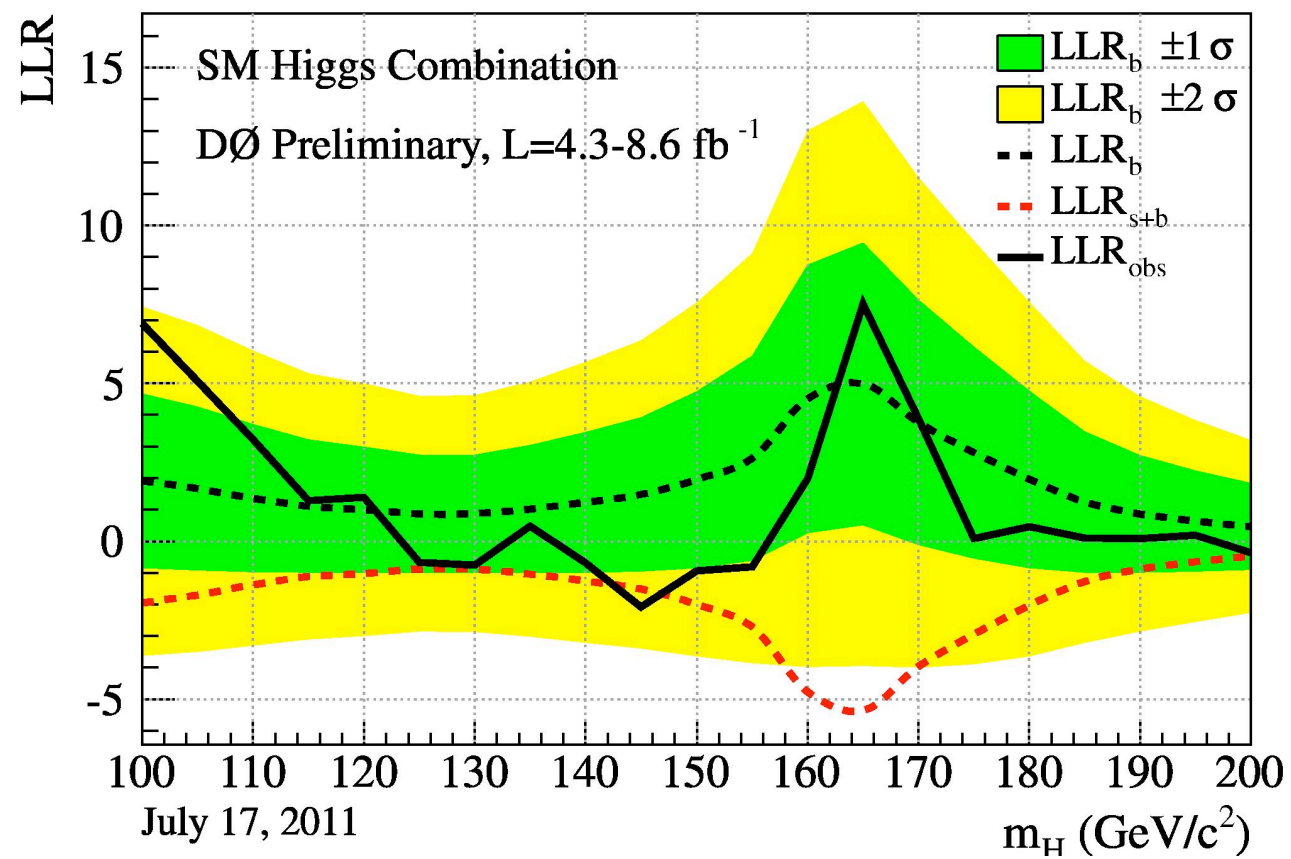
Inputs

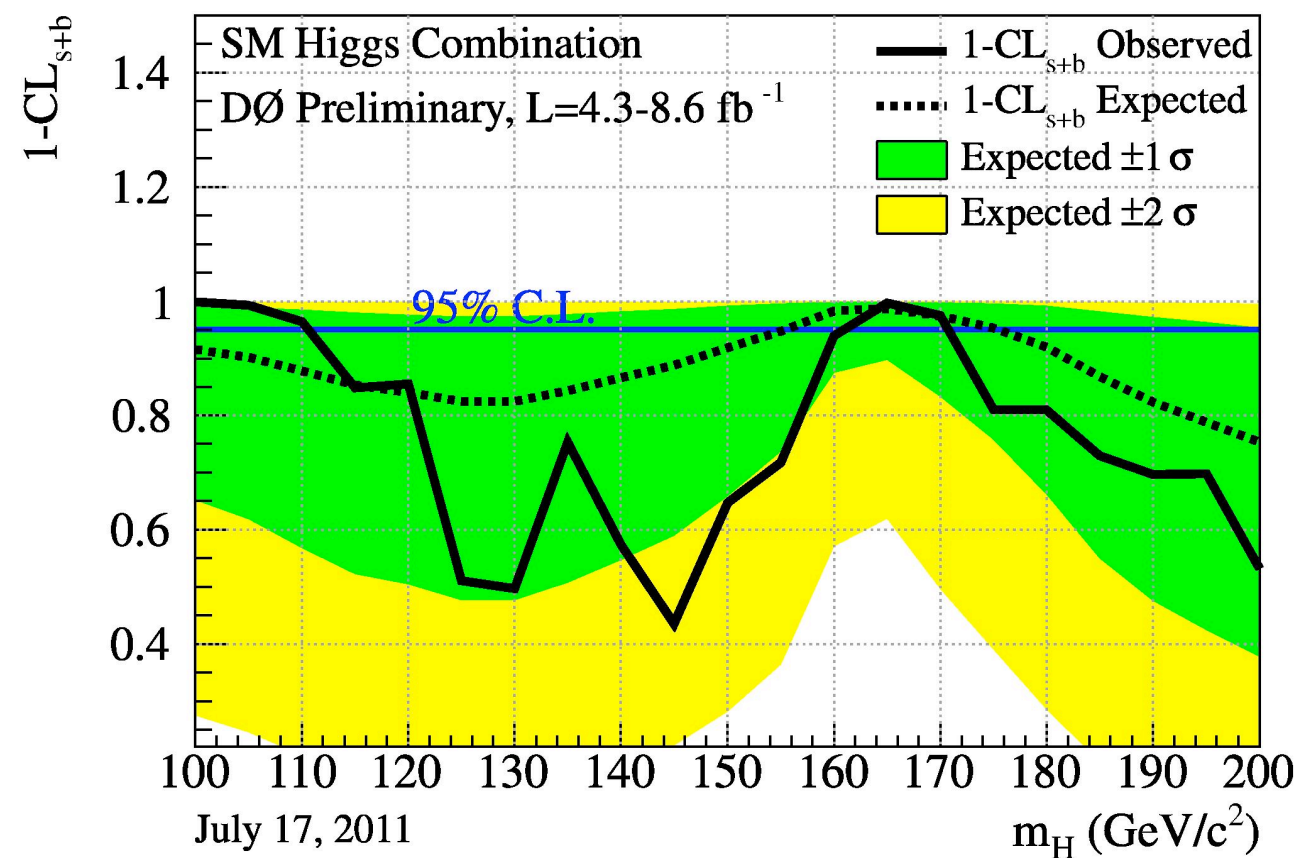
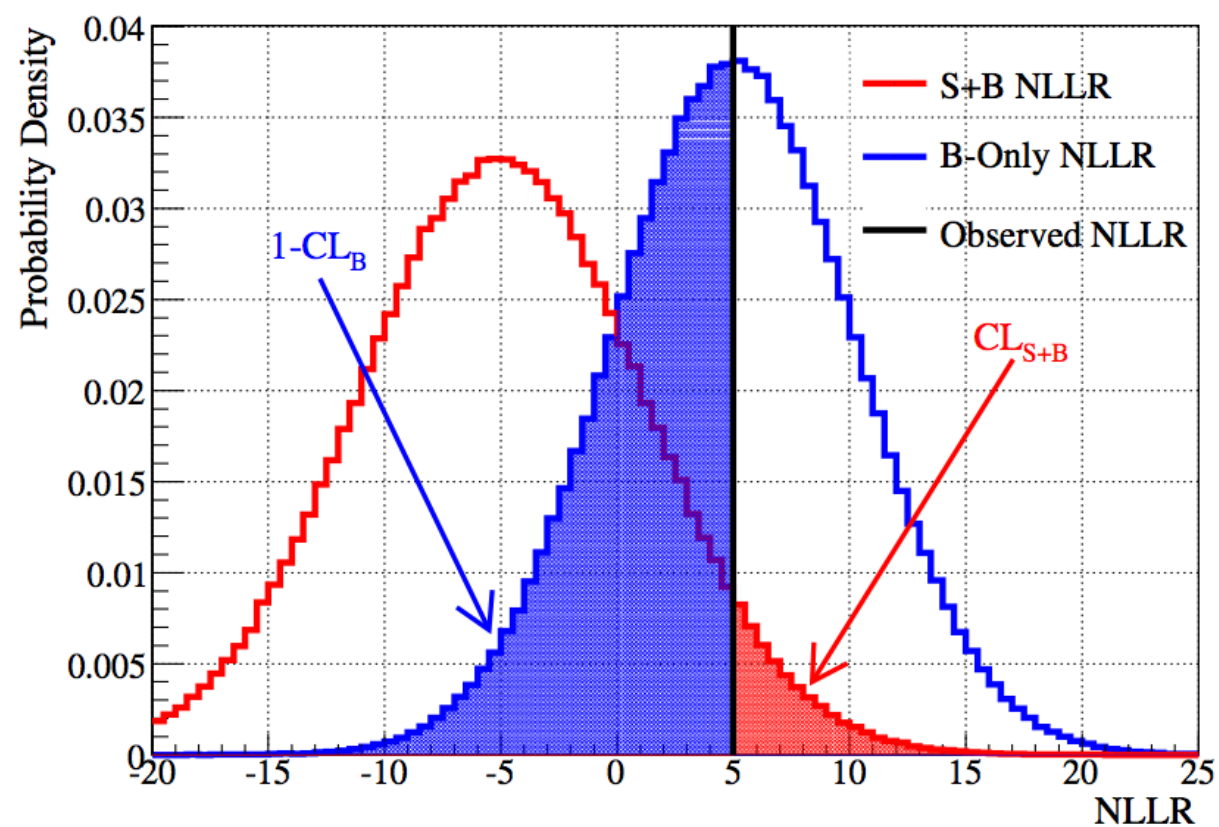


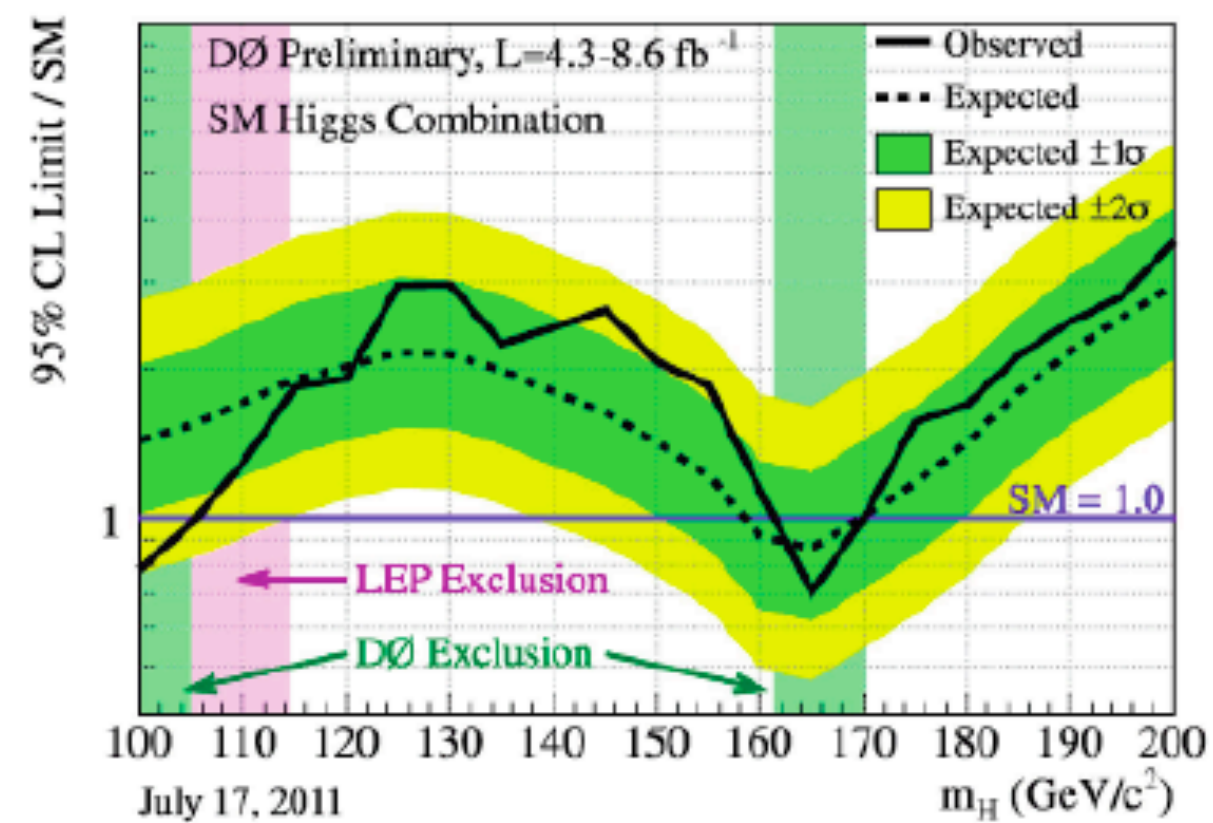
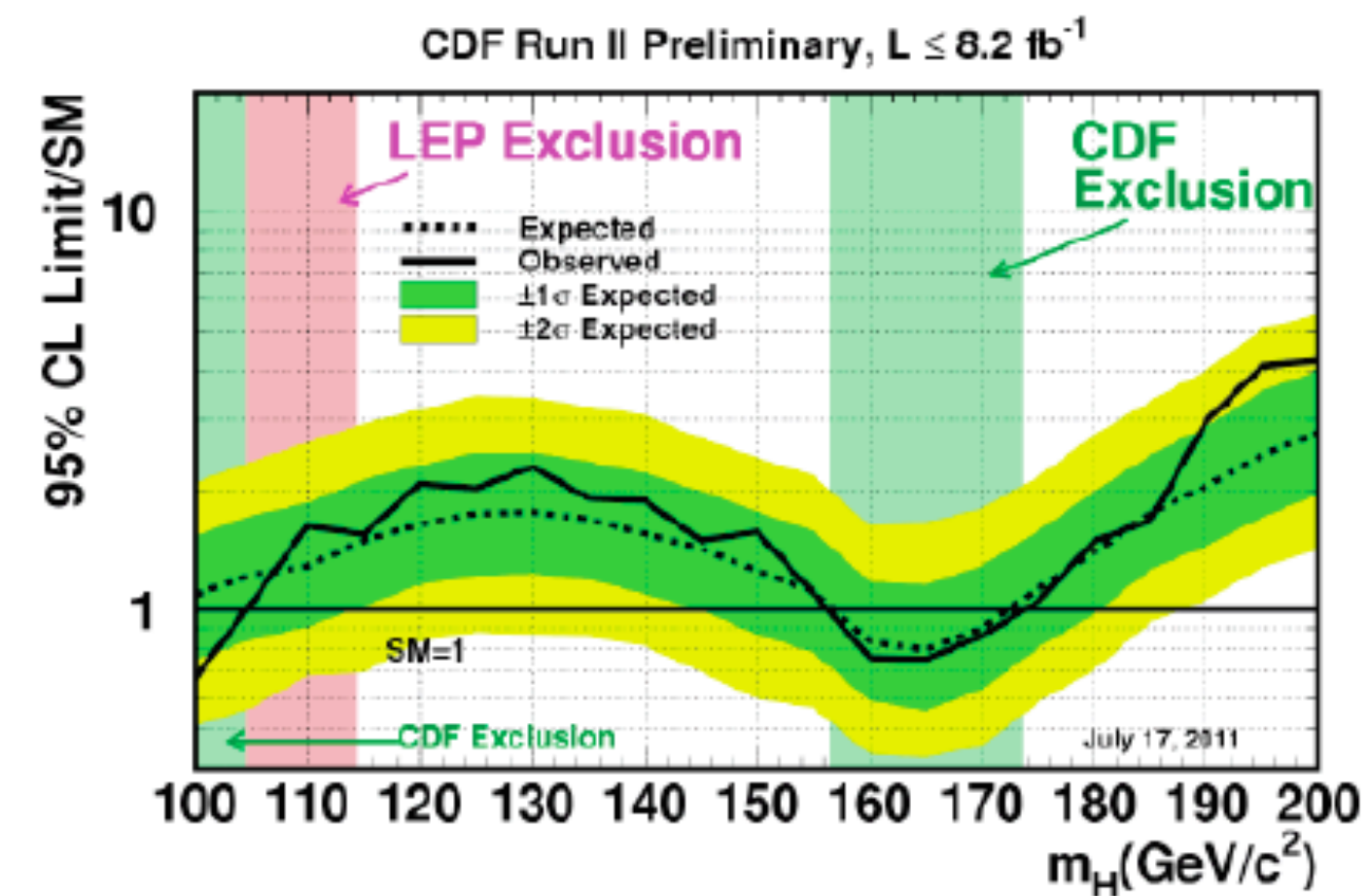
Inputs

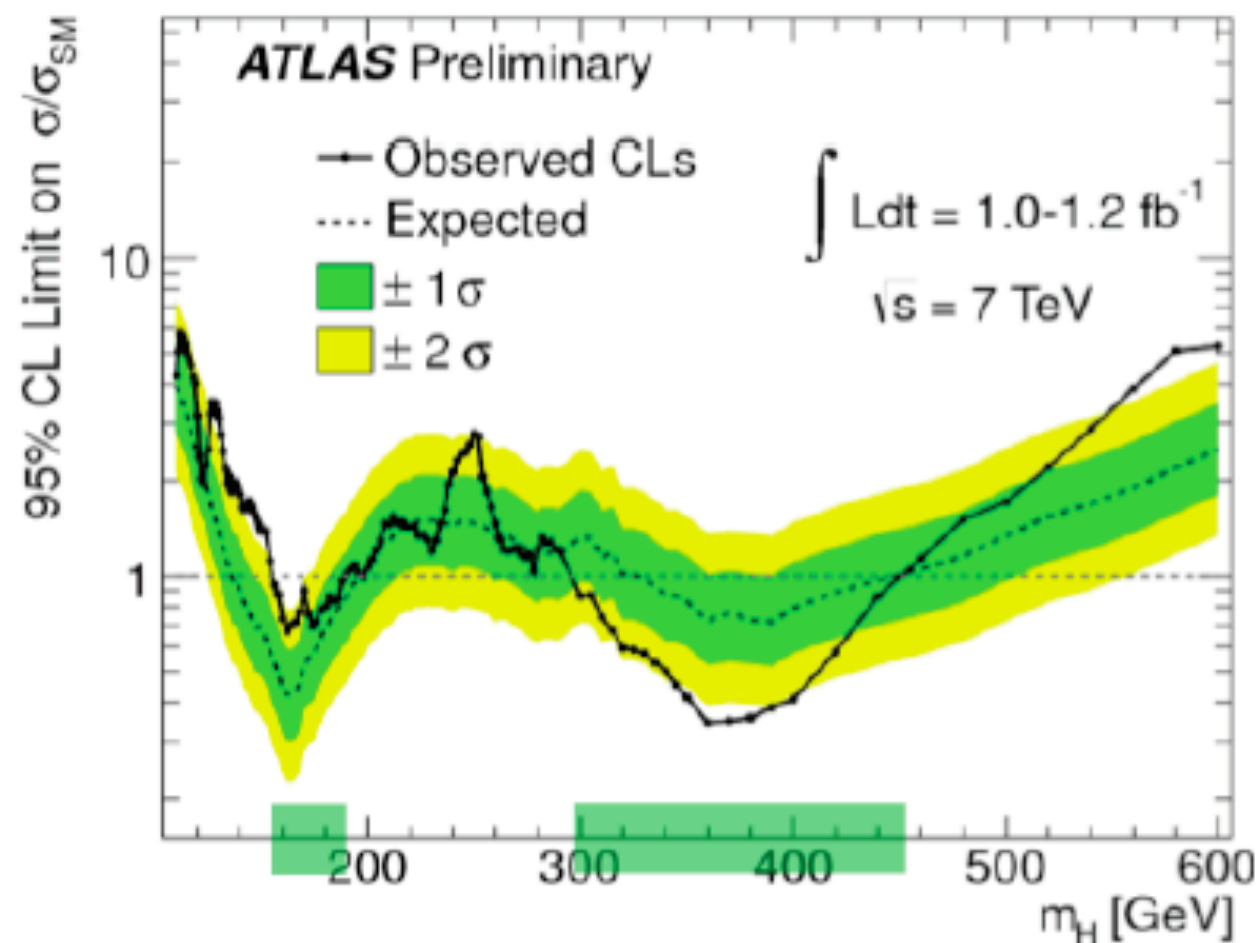


Combining all 9 analyses:



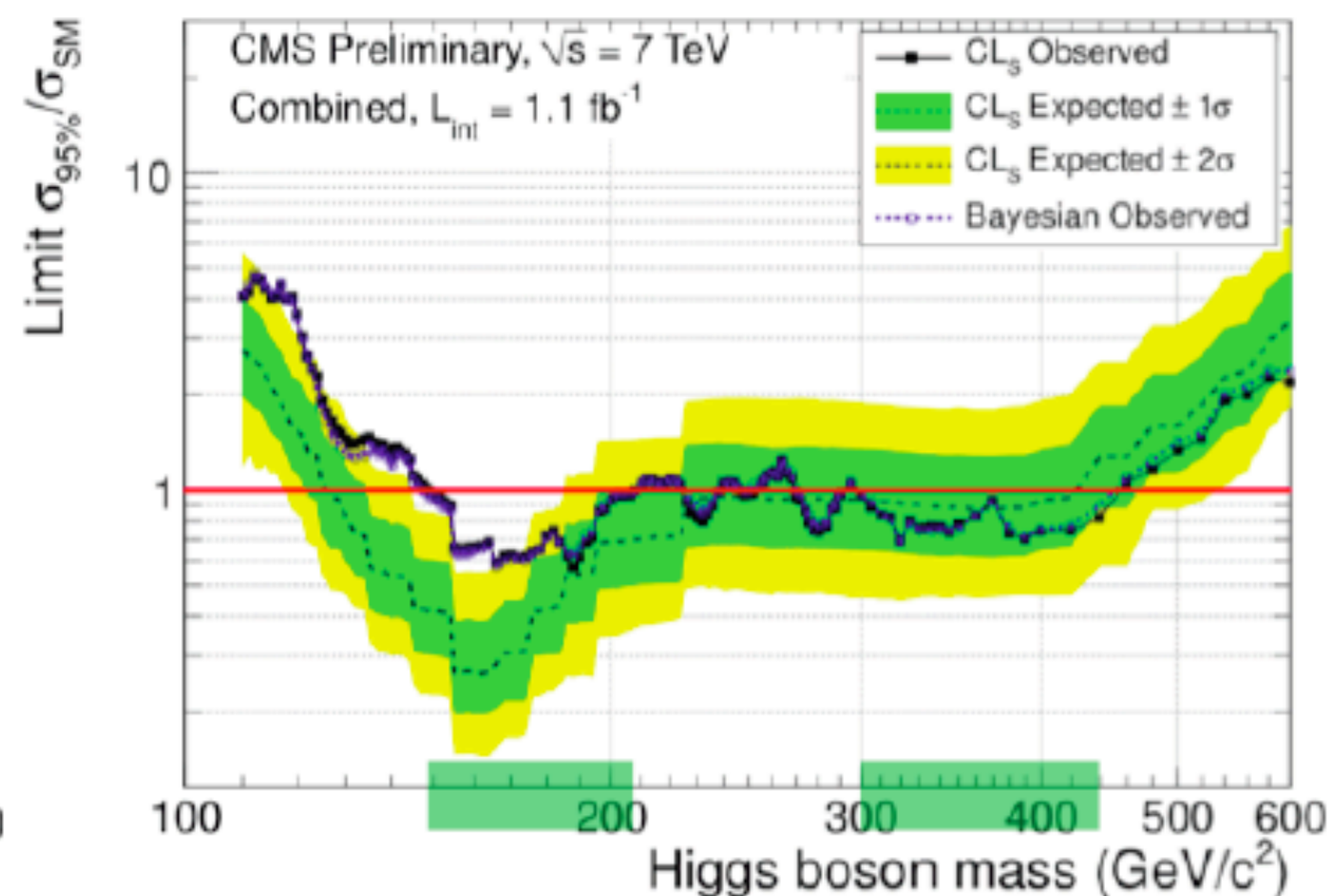






SM Higgs excluded at 95% C.L.

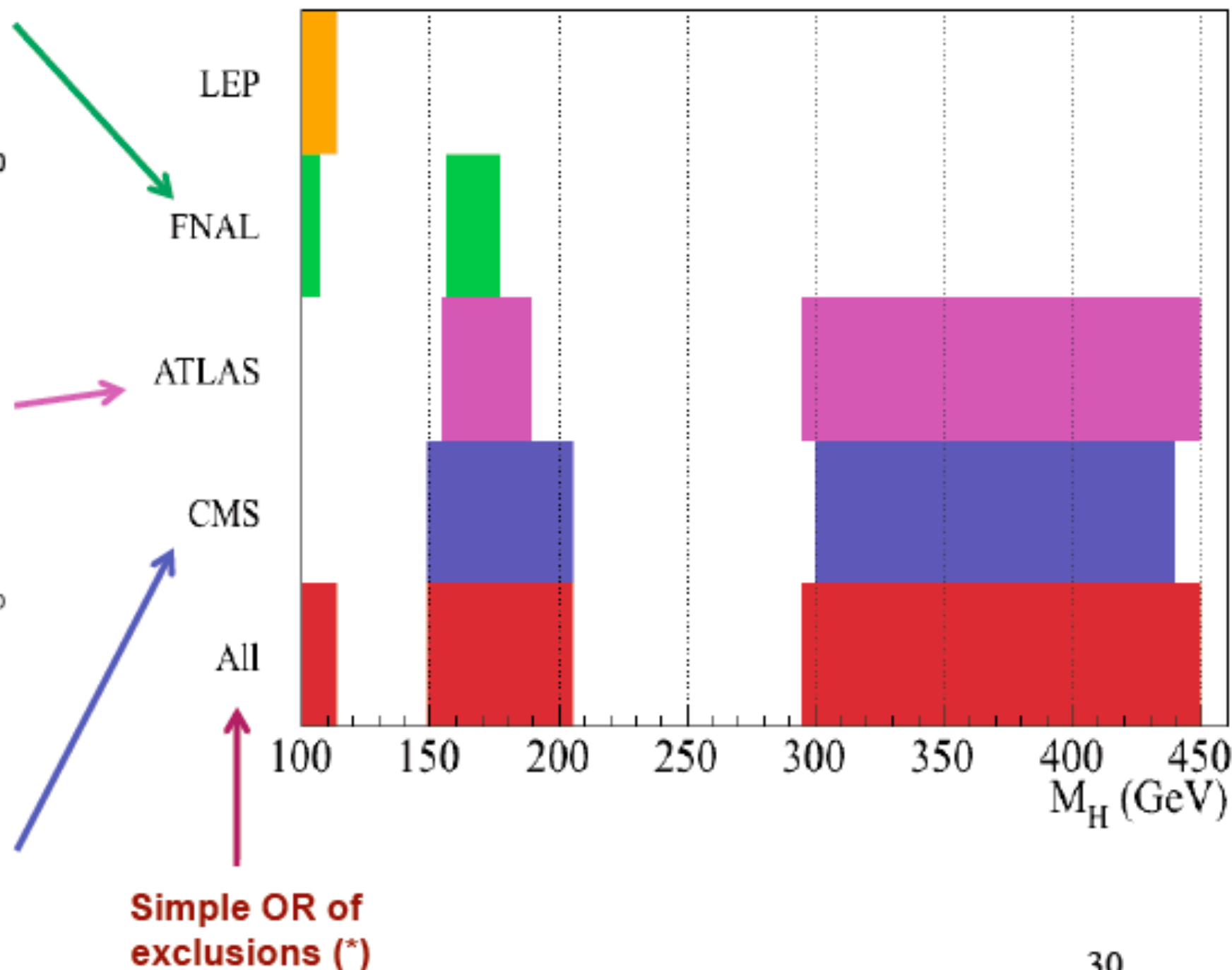
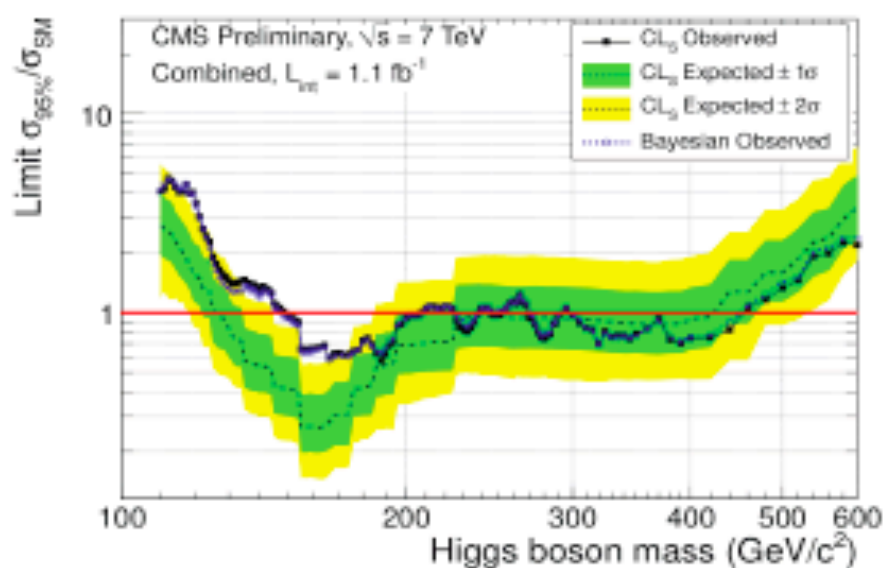
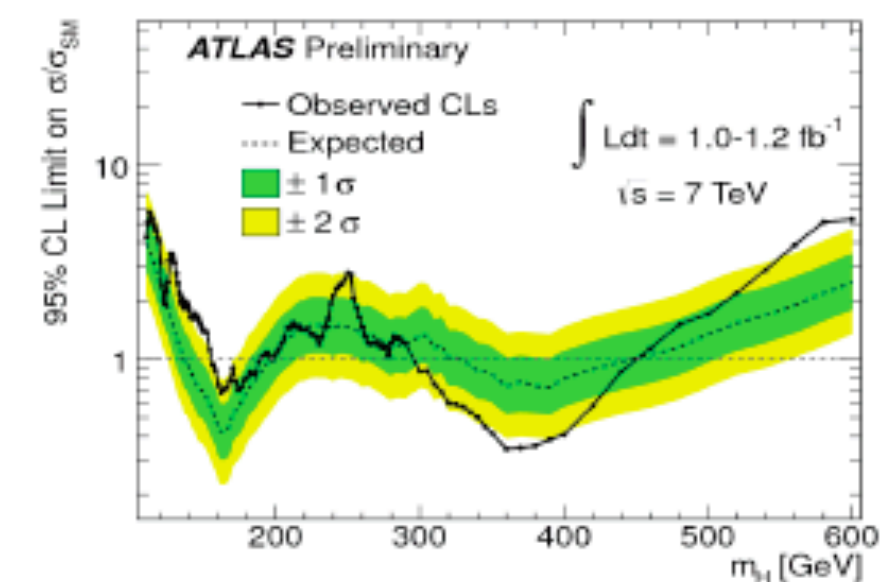
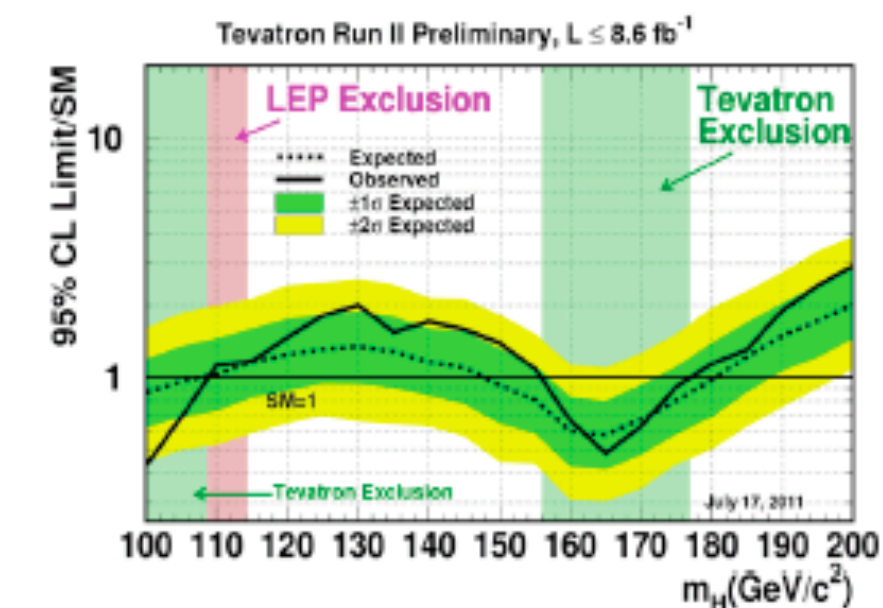
$155 < m_H < 190 \text{ GeV}$ and $295 < m_H < 450 \text{ GeV}$



SM Higgs excluded at 95% C.L.

$149 < m_H < 206 \text{ GeV}$ and $300 < m_H < 440 \text{ GeV}$

Higgs Search Summary



30

(*) Caution, exclusion probability >95%